ANALYSIS OF RELIABILITY OF PARENT-INFANT INTERACTION THROUGH THE USE OF GENERALIZABILITY THEORY

By

WILLIAM G. HUITT

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1978

TO

MY MOTHER AND FATHER

Without whom this project
would not have been attempted

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the assistance of many people. I would first like to express appreciation to my doctoral committee: Dr. Barry Guinagh, for his guidance throughout my graduate program; Dr. Robert Soar, who originally proposed the topic for this dissertation and helped formulate its design; Dr. William Ware, who provided invaluable assistance in the completion of this work; and Dr. Robert Ziller, who has provided assistance in the development of a theoretical base.

Special thanks are due Dr. Ira Gordon who has had such a great influence on my education and career and whose personal friendship I valued highly. His untimely death was a loss to us all.

Special thanks are also due Dr. Maria Llabre who has been helpful in the development of an understanding of generalizability theory. Much of the outline and organization of this dissertation is based on her work.

An expression of appreciation is due the families who consistently came to the studio for tapings, even when it was inconvenient. The staff of the Parent-Infant Transaction Project, Lee Cohen, Vance Geiger, Jan Lederman, Karen Long, Paula Mendelsohn, Pat Schlencker, Ruth Soar and Louise Stephenson proved their dedication time and time again during the year and a half it took to gather the data. Becky Montevideo, Bobbie Wilson and Robin Artis typed the final draft.

A final debt of gratitude must be paid to my best friend and wife, Marsha, for her constant love and support; to Kevin and Geoffrey, who have added much joy to my life; and to the Baha'i Community of Gainesville who provided a much needed opportunity for spiritual growth.

TABLE OF CONTENTS

TABLE OF CONTENTS - CONTINUED

LIST OF TABLES

LIST OF TABLES - CONTINUED

LIST OF TABLES - CONTINUED

LIST OF TABLES - CONTINUED

LIST OF TABLES - CONTINUED

LIST OF TABLES - CONTINUED

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate Council of
the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

ANALYSIS OF RELIABILITY OF PARENT-INFANT INTERACTION
THROUGH THE USE OF GENERALIZABILITY THEORY

By

WILLIAM G. HUITT

December, 1978

Chairman: Dr. Barry J. Guinagh
Cochairman: Dr. Robert S. Soar
Major Department: Foundations of Education

The purpose of this project was to study the reliability of factors of parent-infant interaction through the use of generalizability theory. A review of the literature indicated that coefficients of intercoder agreement were being reported as indications of reliability for measures derived through systematic observation even though two separate studies published 15 years ago had shown that method to be inadequate.

Twenty-eight white, middle-class parents and their first-born infants (14 male, 14 female) were video-taped in a structured teaching situation in a laboratory setting at 19, 25, 37 and 43 weeks of age. Mother and father were taped for three minutes each while engaging their infant in a structured task thought to be appropriate to the age of the infant. Two trained observers coded the video-tapes using the Reciprocal Category System. Thirty-two parent-infant interaction measures accounting for 84 percent of the total interaction tallies

were factor analyzed and the five factors were rotated to the Varimax criterion. Scores were generated based on the resulting five factors and the intercoder agreement and reliability of these scores were analyzed using generalizability theory.

Three separate analyses were done: one for intercoder agreement and two for reliability. One reliability analysis used a design where subjects were crossed with coders and the other used a design where subjects were nested within coders.

The results of the intercoder agreement analysis indicated that four of the five factors showed problems for two occasions. For a third occasion the intercoder agreement coefficients were above .70 for all five factors. The results of the two reliability studies indicated that the value of the generalizability coefficients for two of the five factors were satisfactory when generalization was intended to all parent-infant pairs and all coders, but to only one occasion. A third factor was shown to be moderately satisfactory under the same conditions and two were judged as unsatisfactory. In other words, three of the factor scores of parent-infant interaction were reliable only for mother or father playing with their infant at a specific occasion; two were not satisfactory even under those limited conditions. Observer disagreement and lack of variance between subjects were cited as problems before 37 weeks, and lack of item consistency was thought to be a problem for three of the factors.

The investigator concluded that a design where subjects were nested within coders, used in combination with a design where a portion of the subjects were crossed with coders, provided extensive and valuable

information about the reliability of factors derived through the use of systematic observation. It was demonstrated that researchers can use systematic observation and report detailed reliability data without going to the extra expense of having every subject coded by two observers at every session. Suggestions for future research on parent-infant interaction were outlined.

CHAPTER I

INTRODUCTION

During the past 15 years we have witnessed a significant increase in the number of studies attempting to ascertain the important variables related to infant competence. Lewis (1967) pointed out that this is not a new interest to social scientists, but rather the investigation of old questions by means of more sophisticated techniques, procedures and measures. One of the major changes has been in types of strategies for data collection. As the problems of measures of self-report have been reported in the literature (e.g., Yarrow, 1963), investigators have increasingly turned to direct observation (Clark-Stewart, 1973). As is often the case, change in one aspect of an investigation necessitates that other changes be made as well. For example, reliability of measures is always an important issue and, although a large body of psychometric theory on relability of paper-and-pencil instruments exists, this same issue has only recently begun to be addressed in systematic observation.

Statement of the Problem

Medley and Mitzel (1963) and Cronbach, Rajaratnam and Gleser (1963) have proposed very similar methods of analyzing reliability of observational measures, both of which use ANOVA intraclass correlation coefficients as estimates of reliability. Rowley (1976) has observed that the major advantage of the variance components approach which has

been proposed is that it "enables the researcher to pinpoint multiple sources of error and to compute a number of different reliability coefficients for different purposes." (p. 51) One difference is that Medley and Mitzel (1963) only discuss matched data (every subject rated by all coders) whereas Cronbach, Gleser, Nanda and Rajaratnam (1972) have extended their theory to unmatched ratings (all coders do not rate each subject). This is important since the required extra coding for matched data for an entire project is probably one reason why most researchers are still only reporting intercoder agreement of specific sessions as measures of reliability rather than utilizing a more extensive analysis.

A second reason for the paucity of extensive analysis of reliability in studies using systematic observation is that a user of an observation instrument is not normally concerned with reliability beyond the extent that unreliable data will obscure otherwise valid relationships (Rowley, 1976). He states that when one is speaking of the reliability of a test, "it is usually fairly clear that the term 'reliability' refers to the scores obtained by some sample of examines on that test." (p. 52) The developer of a test is expected to analyze its reliability since reliability is considered to be a property of the instrument itself. Therefore, it is assumed that all investigators will get the same <u>measures</u> if they use a reliable <u>instrument</u> on a sample of the same population.

Rowley (1976) goes on to observe that in the context of systematic observation, "it has frequently been asserted that reliability is a desirable property; it has not always been clear

just what it is that is supposed to possess this attribute." He, therefore, makes the following distinctions:

> Observation instrument - a set of procedures whereby an observer can record and categorize the behavior of a subject or subjects. It normally consists of a number of items, to which the observer responds in some way dependent on the behavior he has observed.
>
> Observation record - a set of data (usually in the form of symbols) which describes the behavior of one or more subjects during one or more periods of observation.
>
> Observation measures - a procedure for using an observation record to assign scores to each of the subjects of observation; each score so assigned is assumed to reflect some characteristic of that subject. (p. 52)

Since single measures are very seldom used by themselves, I would make the following additional distinction:

> Observation composite or factor - a procedure for combining measure scores to assign composite or factor scores to each of the subjects; it is also assumed that these scores reflect some characteristic of the behavior of that subject.

Reliability, then, is a property of a measure, composite or factor score and not of an instrument or record (Rowley, 1976). It is possible that a single observation instrument could produce scores which are reliable and also scores that are unreliable. For example, if one wanted to investigate the relationship of parent verbal behavior to infant competence, one could decide to count specific behaviors such as parent expresses positive affect, asks a question, gives instructions, responds to child's request, corrects child's behavior, etc. The instruction to the coder on what to look for and how to code it (i.e., every time the behavior occurs, once per time period, etc.) would constitute the instrument. When the coder actually observed the parent-infant interaction and wrote something on a code

sheet, an observation record would be produced. From that record several different measures, composites or factors could be produced by counting the frequencies of items individually, by adding certain items together based upon previous literature, by factor analysis, etc. Each measure, composite or factor score is a separate entity and could prove to be either reliable or unreliable.

Medley and Mitzel (1963) have defined reliability of a measure as follows:

> a measure is reliable to the extent that the average difference between two measurements independently obtained in the same classroom is smaller than the average difference between two measures obtained in different classrooms. (p. 250)

By this is meant that if observer A were to code a teacher's classroom on several occasions and observer B were to code the same classroom, though on different occasions, the difference between the average measure would be smaller than if observer B had coded a different classroom.

Using this definition of reliability, unreliability can come about in two ways:

1) Two measures of the same subject (or classroom, parent, etc.) tend to differ too much because:
   a) Behavior of the subject is unstable;
   b) Observers are unable to agree on what occurs;
   c) Different items which enter into the measurement lack consistency, etc.
2) Differences between subjects are too small.

Both item consistency and inter-rater agreement are important issues to be addressed, but they are not the most important sources of error variance. The most important source of error variance is the variability of the behavior of the subject of observation (McGaw, Wardrop and Bunda, 1972), which must be stable for a given subject and must vary among the subjects if reliable scores are to be obtained.

Generalizability Theory

As stated previously Medley and Mitzel (1963) and Cronbach et al. (1963) have published detailed examples on the use of ANOVA as a technique for investigating reliability of systematic observation measures. Cronbach et al. (1963) have labeled their theory generalizability theory which is

> based on the premise that any rating of an individual is only one of a population of ratings that might be made for that individual. The ideal datum would be the average of the population of the ratings, the universe score. The question of reliability is posed as the question of the generalizability of the obtained rating to the universe score. One method of indexing the accuracy of the generalization is a generalizability coefficient, the ratio of the universe score variance to observed-score variance. (Algina, 1978, p. 135)

Cronbach et al. (1972) have distinguished between a generalizability study (G study) which is done for the purpose of investigating the relation between observed score and the universe score, and a decision study (D study) from which decisions are made as to the relationship of the score under study to other measures. Stated another way the purpose of a G study is to investigate the reliability of specific scores which have been derived from an observation record. Each specific score would only be a sample of all universe scores that might be obtained from a specific observation record. A D study could then

use the scores investigated in the G study to investigate those scores' relationship with certain outcome or product variables.

Systematic Observation of Parent-Infant Interaction

Investigators using systematic observation to study parent-infant interaction have not proven to be an exception with respect to reporting intercoder agreement coefficients as estimates of reliability. In a review of 50 observational studies (Lytton, 1971), there were 13 applicable to infancy, none of which reported more than intercoder agreement. In a later compilation of 73 instruments used in child development (Boyer, Simon and Karafin, 1973), 10 were applicable to infancy and none of these reported more than intercoder agreement. From several major publications in the area of development psychology from 1973 to the present (i.e., Child Development, Development Psychology, Monographs of the Society for Research in Child Development), 13 studies of parent-infant interaction were noted and again, none of these reported more than intercoder agreement.

Gordon (1974) investigated mother-infant interaction using three different observation instruments developed in three separate projects (Escalona and Corman, 1974; Gordon and Jester, 1972; Watts and Barnett, 1971). Gordon's study was a direct response to earlier appeals for cross-validation of findings derived from the use of home-grown instruments (Ad Hoc Committee on Child Mental Health, 1971; Sparling and Gallagher, 1971). These same three instruments are presently being used in a longitudinal study of mother- and father-infant interaction in an attempt to extend previously found relationships between parent-infant interaction and infant competence (Gordon and

Soar, in progress). However, intercoder agreement is the most that has ever been reported as estimates of reliability for any scores derived from these instruments.

As important as it is to identify measures, composites or factors of parent-infant interaction through repeated use of observation instruments, it is even more important to investigate repeatedly the reliability of those scores since without this information one does not know if nonsignificant prediction of infant competence is a function of no real relationship or the result of nonreliable data. In addition, the extent to which the measures, composites or factors are generalizable should tell us something about parent-infant interaction separate and apart form whether or not those scores predict infant competence.

Purpose of Study

The purpose of this study, then, is to investigate the reliability of parent-infant interaction through the use of generalizability theory. Important variables which have been found to relate to either parent or infant behavior will be used to define the different facets of the study. Several different designs will be used in an attempt to determine the feasibility of eliminating the requirement to use matched data when investigating reliability of data which are to be used in a decision study. Since this study may be some readers' first introduction to generalizability theory, a Glossary is provided. It is expected that the results of this study will contribute to the theory of parent-infant interaction as

well as the design of future decision studies relating parent-infant interaction to infant competence.

CHAPTER II

REVIEW OF THE LITERATURE

There are three major areas of research to be reviewed. First is the literature on parent-infant interaction; second is the literature on systematic observation; and third is the literature on generalizability theory.

Parent-Infant Interaction

Most of the research that has been done on parent-infant interaction during the first year of life has been limited to the study of the mother-infant dyad. The exclusive study of this dyad assumes that it is unique (Ainsworth, 1973; Bowlby, 1951; Stern, 1974) and the forerunner of later social relationships (Kogan, Wimberger and Bobbitt, 1969); the father-infant relationship has been severely neglected (Lamb, 1975). However, recent research has indicated that the father-infant relationship is established early and contributes significantly to the development of the child.

In one recent study, Lamb (1977) observed 20 infants (10 boys, 10 girls) interacting with their mothers and fathers at home when they were 7, 8, 12, and 13 months of age. The families were described as a

> representative sample of young, intact, and stable lower- to upper-middle-class families within which parental and marital roles were traditionally allocated. (p. 169)

The major variables of interest were affiliative behaviors (smiling, vocalizing, looking, laughing and proferring) and attachment

behaviors (proximity, touching, approaching, seeking to be held, fussing and reaching). Lamb found that infants showed no clear preference for either parent in the display of attachment behaviors, even though both parents were consistently differentiated from a relatively unfamiliar investigator on these measures. This was taken as an indication that infants were clearly attached to both parents at a very early age. Preference in the display of affiliative behaviors were explained largely by differences in the degree of adult activity in interaction with the infants. As the infants grew older, they were increasingly likely to direct affiliative behaviors to both the parents and the investigator.

Additional evidence that father-infant interaction should be studied is provided by a review of four different experimental studies of infant attachment to both mother and father (Kotelchuck, 1976). He reported that even though the exact number of children who prefer mothers or fathers varies, depending on the measure chosen, approximately 55% of the 12- to 21-month-old children showed maternal preferences, 20% showed joint preferences, and 25% showed paternal preferences. The data from these studies were gathered in the laboratory, in the home and cross-culturally. In view of this research, Bowlby's (1969) conclusion that the infant is monotropically (relates to one person) matricentric (relates to mother) seems to need some modification.

Additional support that the father should be included in patterns of studies of parent-infant interaction comes from the fact that <u>sex of parent</u> is one of the most consistently reported variables that has been found to influence these patterns. In a study of 24 4- and 8-month-old infants in 11 Israeli kibbutzim, Gewirtz and Gewirtz (1968) found that the infant

sees his/her mother for at least twice as much time as he/she does the father. However, when the type of situation was divided into "caretaking" and "social," it was found that the differences were accounted for by the fact that fathers spent very little time in caretaking activities.

Rebelsky and Hanks (1971), in a study of 10 2-week to 3-month-old infants (7 male, 3 female) born into white, lower-middle- to upper-middle-class families, found that fathers talk infrequently and for short periods of time to their infants during the first three months of life. Also, fathers tend to vocalize differently with their male and female infants, which when compared to similar data for mothers (Moss, 1967), suggests that it is just the opposite of the mother-infant pattern. Rebelsky and Hanks's data indicate that at 2 weeks and 4 weeks of age fathers of female infants tend to verbalize more while Moss's data show that mothers of male infants verbalize more at 3 weeks. By the time the infants are three months of age, these patterns are reversed. Fathers of male infants vocalize somewhat more whereas mothers of female infants tend to vocalize more.

Additional evidence that mothers and fathers differ in their interactions with infants is provided by several additional studies. An informal study by Biller (1974) as reported in Lamb (1976c) suggests that whereas mothers were more likely to inhibit a child's exploration, fathers encouraged their infant's curiosity and urged them to attempt to solve cognitive and motoric challenges.

Lamb (1976d) found that mothers held their infants most often to engage in caretaking functions, while fathers held them most often to play. Kotelchuck (1976) also found that fathers spent a larger percentage of their time in playful behavior with their infants than did mothers.

The major variables that have been found to correlate consistently with infant behavior are age and sex of the infant. Kagan (1971), in a study of 180 white, firstborn infants, found discontinuities in behavior when assessing infants on a wide variety of measures at 4, 8, 13, and 27 months. Gordon and Jester (1972) found very similar discontinuities with a sample of black infants born into low-income families. Emde, Gaensbauer and Harmon (1976) found that changes in unexplained fussiness, wakefulness, and the emergence of infant social behaviors combined to distinguish three separate periods during the infant's first year of life: 1) Birth to the end of the second month, 2) third month through the sixth month, and 3) seventh month through the twelfth month. Bell and Harper (1977), in a review of the literature of discontinuity in infancy, concurred with these findings, adding that the increased ability of the infant to react to the stimulation of the environment is also a basis for change.

These discontinuities in behavior have been replicated in cross-cultural research by Lusk and Lewis (1972) who studied 10 Wolof infants in Senegal. That these discontinuities result in altered mother-infant behavior patterns was shown by Crawley, Rodgers, Freidman, Iacobbo, Criticos, Richardson and Thompson (1978) who studied 48 4-, 6-, and 8-month-old infants and their mothers in a laboratory free-play situation. Mothers

of 4-months-old typically played games conducive to the direct stimulation of infant attentional and positive affect responses. In contrast, mothers of 8-month-olds incorporated games possessing a conventional motoric role that could readily be assumed by her infant.

With regard to infant sex, Clark-Stewart (1973), in a longitudinal study of 36 low-income mothers and their first-born, normal infants (9-18 months old), found that boys became increasingly more object oriented with age whereas girls became increasingly more socially oriented. Goldberg and Lewis (1969) observed 64 13-month-old infants (32 male, 32 female) with their mothers in a standardized laboratory free-play situation and found striking sex differences. Boys were more independent, showed more exploratory behavior, played with toys requiring gross motor activity, were more vigorous, and tended to run and bang in their play. Girls were more dependent, showed less exploratory behavior, and their play reflected a more quiet style. Moss (1967) also found that at age three months boys slept less and cried more than girls.

Type of situation is another variable which has been found to influence parent-infant interaction. Rebelsky and Hanks (1971) found that fathers decreased their verbalizations with their infants during caretaking activities. Lamb (1976a) showed that infants' affiliative behaviors towards their parents changed when a stranger entered the room. Clark-Stewart (1973) found that parent-infant interaction patterns were different in free-play and structured situations. However, Peterson (1975), in a study of 20 white, middle- and working-class mothers and their 12- to 16-month-old infants, found many categories of interactive behavior similar from home to a laboratory setting.

Several parent behaviors which have been found to contribute to infant competence are stimulation, responsiveness, acceptance of child's behavior and appropriateness of parents' behavior for age and ability of the child (Bell and Ainsworth, 1972; Bromwich, 1976; Clark-Stewart, 1973; Gordon, 1974; Kagan, 1971; Lewis and Goldberg, 1969). Bakow, Sameroff, Kelly and Zax (1973) and Clark-Stewart (1973) have shown that these parent behaviors tend to cluster together; parents scoring high on one generally score high on all. Ability to sustain interaction sequences (Peterson, 1975; Watson, 1972), amount of language and positive feedback produced (Peterson, 1975) and the ability to provide a warm, nuturant atmosphere (Lamb, 1976b; Walters and Stinnett, 1971) have also been found to relate positively to infant competence.

Systematic Observation

In order to study parent-infant interaction properly one must observe the parent and infant together and record the behavior in such a manner that the sequences of behavior are preserved for later analysis (Gerwirtz, 1969). However, the vast majority of researchers studying parent-infant interaction have used time-sampling observation systems in which the sequencing of events is not recorded. In a review of 50 observational studies (Lytton, 1971), there were 13 applicable to infancy and only one study which preserved the sequences for later analysis (Brody, 1956). In a later compilation of 73 observation instruments used in child development (Boyer, Simon, and Karafin, 1973), 10 were applicable to infancy. Of those, only three preserved the parent-infant sequences for later analysis (the systems of Ainsworth, Salter, Bell, and Stayton, 1972; Caldwell and Honig, 1970; Gordon and Jester, 1972).

The Ainsworth, Salter, Bell, and Stayton (1972) system was intended for use in coding infant attachment and reciprocal maternal behaviors during naturalistic observation in the familar home environment. The Caldwell and Honig (1970) system was also designed to be used in natural settings and provides data on behavior emitted and received by children in relation to their peers, their caretakers, their teachers and the environment. Both of the above named systems code both verbal and nonverbal behavior. The Gordon and Jester (1972) system was designed to code the interaction of a mother, an infant, and a parent educator in a structured teaching situation. In the Gordon and Jester system adult behavior must be verbal although baby behavior may be either verbal or nonverbal.

Generalizability Theory

The organization of this section was influenced by Llabre (1978). Classical psychometric theory is primarily based on a model proposed by Spearman (1904) which states that a person's observed score (X) is the sum of two components, one being the true score (T) and the second being undifferentiated error component (E) as shown below:

$$X = T + E.$$

Since these two components are assumed to be independent of each other, the variance for a group of individuals can be partitioned into the sum of independent variance components:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2.$$

General agreement has been reached that the form of the definition of reliability should be a ratio of true score to observed score

$$r_{XX} = \frac{\sigma_T^2}{\sigma_X^2},$$

but there is still some disagreement as to how σ_T^2 and σ_X^2 should be computed (Medley and Mitzel, 1963).

With respect to the use of generalizability theory for the purpose of estimating the reliability of scores obtained through systematic observation, the underlying linear model is

$$X_{pr} = \mu + (\mu_p - \mu) + (\mu_r - \mu) + e_{pr},$$

where X_{pr} is the rating of the *p*th subject by the *r*th rater (Algina, 1978). Additionally,

> Both raters and subjects are assumed to be random samples from infinite populations of raters and subjects. Using $\varepsilon(\quad)$ to represent the expectation operator, the various μs are defined as
>
> $$\mu_p = \underset{r}{\varepsilon} X_{pr},$$
>
> $$\mu_r = \underset{p}{\varepsilon} X_{pr},$$
>
> $$\mu = \underset{rp}{\varepsilon\varepsilon} X_{pr}$$
>
> The mean μ_p as the expected value of all ratings of the *p*th subject is the quantity we would like to obtain for this subject and is referred to as the universe score. The residual e_{pr} is composed of two confounded components,
>
> $$e_{pr} = \alpha_{pr} + E_{pr},$$
>
> where α_{pr} is the interaction between the *r*th rater and the *p*th subject, and E_{pr} is an error random variable with mean zero and variance σ_{pr}^2. (Algina, 1978, p. 136)

The variance σ_p^2 is referred to as the universe score variance (or true score variance) and is equal to

$$\sigma_p^2 = \underset{p}{\varepsilon}(\mu_p - \mu)^2 .$$

The variance for raters is

$$\sigma_r^2 = \underset{r}{\varepsilon}(\mu_r - \mu)^2 ,$$

and the error variance is

$$\sigma_E^2 = \underset{rp}{\varepsilon\varepsilon}(\sigma_{pr}^2 + \alpha_{pr}^2).$$

In the case of generalizability theory, then, the error variance (σ_E^2) is composed of variance for raters (σ_r^2) and error variance (σ_e^2):

$$\sigma_E^2 = \sigma_r^2 + \sigma_e^2 .$$

There are two separate definitions of observed-score variance derivable from the above definitions and equations. The first is

$$\sigma_{x|r}^2 = \underset{p}{\varepsilon}(X_{pr} - \mu_r)^2 = \sigma_p^2 + \underset{p}{\varepsilon}(\sigma_{pr}^2 + \alpha_{pr}^2)$$

and the second is

$$\sigma_x^2 = \underset{pr}{\varepsilon\varepsilon}(X_{pr} - \mu)^2 = \sigma_p^2 + \sigma_r^2 + \sigma_e^2 .$$

Since the variance for raters (σ_r^2) is very difficult to compute, Cronbach et al. (1963, 1972) defined a generalizability coefficient (ρ^2) as follows:

$$\rho_{p\cdot}^2 = \sigma_p^2 \,|\, \underset{r}{\varepsilon}\sigma_{x|r}^2 = \sigma_p^2/(\sigma_p^2 + \sigma_e^2),$$

which Rajaratnam, Cronbach and Gleser (1965) showed to be a lower bound for $\varepsilon\rho_r^2$.

As observed by Llabre (1978), Brennan (1975) extended the rationale of estimating reliability through the use of generalizability theory to a split-plot factorial design where students were nested within

classes. He compared the generalizability coefficients derived from the split-plot design to coefficients derived when the nesting classification was ignored (i.e., a randomized block design) and found that

> if one uses a randomized block design to calculate reliability for persons when, in fact, persons are nested within some dimension, such as schools or classrooms, the resulting coefficient will be biased, and, moreover, the direction of bias will be unknown. (Brennan, 1975, p. 785)

As also noted by Llabre (1978), Kane and Brennan (1977) extended the use of the split-plot design to a mixed model (i.e., a model having both random and fixed facets). They showed that different coefficients could be generated depending upon the definition of universe score and, hence, a different definition of error. In each case, however, the definition of observed score was identical. This observed score was simply the sum of the variances of the separate effects, eliminating any effects that were common to all classes or subjects.

Summary

In the review of the literature for parent-infant interaction it was shown that there are four facets which have been found to influence parent-infant interaction. They are <u>sex of parent</u>, <u>sex of infant</u>, <u>age of infant</u>, and <u>type of situation or task</u>. Additionally, the facet of <u>coder</u> has been shown to be an important consideration when using systematic observation.

The review of literature for systematic observation has shown that even though the coding of sequences of behavior is necessary in order to properly study parent-infant interaction, as of the last major review (Boyer, Simon and Karafin, 1973) only three systems had done so

(the systems of Ainsworth, Salter, Bell and Stayton, 1972; Caldwell and Honig, 1970; Gordon and Jester, 1972). Additionally, none of these investigators have reported more than intercoder agreement as measures of reliability.

The review of the literature for generalizability theory has shown that the classical reliability coefficient defined as the ratio of true score variance (σ_T^2) to observed score variance (σ_X^2)

$$r_{XX} = \frac{\sigma_T^2}{\sigma_X^2},$$

is defined in generalizability theory as the ratio of universe score variance (σ_p^2) to observed score variance (σ_X^2), and can be estimated through computation of a generalizability coefficient

$$\rho_p^2 = \sigma_p^2/(\sigma_p^2 + \sigma_e^2)$$

where ρ_p^2 is the generalizability coefficient for the rating of a subject, σ_p^2 is the variance component for the subjects, and σ_e^2 is the variance component for the experimental error associated with that rating. Additionally, it was stated that the use of ANOVA is the accepted method of computing estimates of variance components.

CHAPTER III

DESIGN

The purpose of this study was to investigate the reliability of parent-infant interaction through the use of generalizability theory. Video-tapes of parents interacting with their infants in a structured teaching situation in a laboratory setting were made on four separate occasions. The sample, specific objectives, and data gathering and analysis procedures are described in this chapter.

Sample

The sample consisted of 28 white, first-born infants falling within the normal physical range as determined by physical examinations at age three months. There were 14 male and 14 female infants. This sample was recruited for a larger, longitudinal project (Gordon and Soar, in progress) via local radio, community and campus newspapers, commercial and public television and local pediatricians. Socio-economic status was determined by using the Two Factor Index of Social Position (Hollingshead, 1957). The numbers of families in each of the classifications are shown in Table 1.

Procedure

Each of the 28 families was video-taped for seven observations scheduled six weeks apart beginning at 13 weeks and terminating at 49 weeks. Only the data from the 19 week, 25 week, 37 week and 43 week visits were used in this analysis. At each visit the parents were

Table 1

Number of Families for Each Classification of Hollingshead's Index of Social Position

I	Major Professionals and Executives	6
II	Lesser Professionals and Managers	9
III	Clerical, Sales and Technical	12
IV	Skilled Manual Employees	1

presented with a specific task taken from Gordon's (1970) Baby Learning Through Baby Play, thought to be appropriate for the age of the child (see Appendix). All families at each age were presented with the same activity. The parents were asked to read the instructions and to then use the ideas in interacting with the baby.

Each session consisted of a total of nine minutes: three minutes were mother-infant, three minutes were father-infant, and three minutes were both parents-infant. The order of parents interacting with their infants was randomly assigned. Only the mother-infant and father-infant data were used in this study.

The video-taping took place in a studio on the University of Florida campus that had been temporarily set-up for this purpose. It consisted of a 12' x 12' enclosure with small holes cut in several panels, three black-and-white cameras, monitors, camera mixer, and two 1/2" reel-to-reel Sony recorders.

The Bayley Mental Development Quotient (Bayley, 1969) was administered at age 12 months.

Specific Objectives

The specific objectives of this study were to determine the extent to which factor scores of parent-infant interaction are generalizable across different levels of the facets which have been proposed as affecting those scores.

Objective 1: To determine the intercoder agreement for factor scores of parent-infant interaction at particular occasions.

Objective 2: To determine the reliability of factor scores of parent-infant interaction when subjects are crossed with coders (i.e.,

all subjects are observed by both coders).

Objective 3: To determine the reliability of factor scores of parent-infant interaction when subjects are nested within coders (i.e., each coder observes only one group of subjects).

Method of Analysis

Observer Training

Two observers were trained to code the Reciprocal Category System (RCS) until the measures of a single family on each of the separate items in the observation instrument were within two tallies of each other. Parent-infant interaction measures for the RCS are normally tallied in such a manner that each entry in the record is paired with the entry immediately following it. This produces a 28 x 28 matrix, of which most entries are zero. Since at the time of the training of the coders it was not known which interaction measures would be used to form factor scores nor in what manner the factor scores would be formed, it simply was not feasible to work with anything more than the separate items. The check on the training procedure was simply an attempt to produce an accurate observation record from which observation measures could be produced. Periodic checking was done to insure that the coders did not "drift" from their original agreements.

RCS Measures

The Reciprocal Category System (RCS) which is being used in the present study is a slight modification of that used by Gordon and Jester (1972). It traces its history back through Ober, Wood and Roberts (1968), and Flanders (1965) to Bales' Interaction Process Analysis (1951). The RCS consists of 28 categories, the first digit of which signifies the

actor (infant, mother, father) and the second digit signifies the behavior (see Table 2). Adult behavior must be verbal in nature, but infant behavior may be either physical or verbal. Behavior is coded as it occurs with a minimum of one code every three seconds.

From an analysis of data gathered in a previous project (Gordon, 1974) and a preliminary review of some of the data, 32 measures accounting for 84% of the total tallies were identified for use in further analysis (see Table 3). It was assumed that the raw data would not meet the restricted assumptions of the statistical procedures to be used in latter analysis. Therefore, the data were area transformed (making the distribution as normal as possible) and t-scored measures were produced having a mean of 50 and a standard deviation of approximately 10. The measures were then factor analyzed and rotated to the Varimax criterion using a factor analysis program from the University of Florida Educational Evaluation Library as modified by L. B. Stebbins. Incomplete factor scores were produced for each rotated factor by weighting each variable either one or zero with a cutoff of .3999 (see Horn, 1965). Any measure which loaded above the cutoff on two factors was loaded on both.

Intercoder Agreement

An intercoder agreement analysis was done for three separate occasions and was the method of analysis for Objective 1. At 25 weeks there were 14 families (6 boys, 8 girls); at 37 weeks there were 18 families (12 boys, 6 girls); at 43 weeks there were 17 families (11 boys, 6 girls). Only six families were coded at all three sessions.

Each session was analyzed separately using a randomized block

Table 2

Reciprocal Category System
Description of Behavior

Baby[a]		
01	WARMS	- Baby smiles, laughs, gurgles, cooes, etc. Self-reinforcing behavior such as thumb sucking is also warming behavior.
02	ACCEPTS	- Passive acceptance of situation; the child does not ignore the person or object but neither does he respond.
03	AMPLIFIES	- Simple imitation or expansion of behavior which is begun by parent.
04	ELICITS	- Attempts to get parents to respond which are made in a questionning manner; "Asks" for help or assistance.
05	RESPONDS	- Baby responds appropriately (may be correct or incorrect) to parent eliciting, initiating or directing.
06	INITIATES	- Exploratory behavior which has no observable antecedent.
07	DIRECTS	- Any behavior which attempts to get adult attention or direct adult activity.
08	CORRECTS	- Task-related ignoring behavior.
09	COOLS	- Task irrelevant, emotional expressions; Active or passive agression, crying, hitting, biting, being uncooperative, etc.
10	SILENCE	- Pauses, periods of no activity.
	CONFUSION	- Yawns, sneezing, coughing, wetting, "accidents" or unintentional interruptions; Period of confusion in which communication cannot be understood by observer.
	BABY SLEEPING	- Baby goes to sleep.

[a]may be either verbal or non-verbal in nature.

Note: Adapted from Reciprocal Category System For Use In the Parent-Infant Transaction Project by I. J. Gordon with J. C. Lederman and W. G. Huitt. NIMH # 1 ROL MHIHD 27480-DI, 1976.

Table 2 - continued

Mother[b]			Father[b]
11	WARMS	- Tends to reduce or release tension and/or alleviate threat. Cooing, laughing, clarifying and accepting the feelings and emotions of another are specific examples. Encourages or praises in non-task oriented behavior. Deals mainly with socioemotional climate.	21
12	ACCEPTS	- Positively reinforces or accepts task related behavior of another.	22
13	AMPLIFIES	- Clarification of, building on, and/or developing of actions, behaviors, comments, and/or ideas.	23
14	ELICITS	- Asks questions or requests information about content, subject, or procedures being considered, with the intent that the other should answer and respond appropriately.	24
15	RESPONDS	- Gives answers or responds to questions or requests for information that are directed, initiated or elicited by another person.	25
16	INITIATES	- Statements of facts, information, and/or opinions and ideas concerning the content, subject, or procedures being considered which are self-initiated.	26
17	DIRECTS	- Giving of directions, instructions, orders and/or assignments to which another is expected to reply.	27
18	CORRECTS	- Task-related behavior which tells another that the answer or behavior of another is inappropriate or incorrect.	28
19	COOLS	- Non-task related behaviors which tend to create tension; implied are efforts toward sarcasm, ridicule, regimentation, or alienation of another (i.e., bawling out someone, refecting or criticizing the opinion or judgments of another, or excercising control in order to gain or maintain authority in situation).	29

[b] must be verbal in nature.

Table 3

Key RCS Variables from
Parent-Infant Transaction Project

---

1 Baby warms, accepts; baby warms, accepts
Rows 1,2; Col. 1,2

2 Baby warms, accepts; baby amplifies
Rows 1,2; Col. 3

3 Baby warms, accepts; baby responds
Rows 1,2; Col. 5

4 Baby warms, accepts; baby initiates
Rows 1,2; Col. 6

5 Baby warms, accepts; parent accepts, amplifies
Rows 1, 2; Col. 12,13

6 Baby warms, accepts; parent elicits, initiates, directs
Rows 1,2; Col. 14,16,17

7 Baby ampilifies; baby warms, accepts
Row 3; Col. 1,2

8 Baby amplifies; baby amplifies
Row 3; Col. 3

9 Baby amplifies; baby responds
Row 3; Col. 5

10 Baby amplifies; baby inititates
Row 3; Col. 6

11 Baby amplifies; parent accepts, amplifies
Row 3; Col. 12,13

12 Baby amplifies; parent elicits, initiates, directs
Row 3; Col. 14,16,17

13 Baby responds; baby warms, accepts
Row 5; Col. 1,2

14 Baby responds; baby amplifies
Row 5; Col. 3

15 Baby responds; baby responds
Row 5; Col. 5

16 Baby responds; baby initiates
Row 5; Col. 6

Table 3 - continued

17 Baby responds; parent accepts, amplifies
Row 5; Col. 12,13

18 Baby responds; parent elicits, initiates, directs
Row 5; Col. 14,16,17

19 Baby initiates; baby warms, accepts
Row 6; Col. 1,2

20 Baby initiates; baby amplifies
Row 6; Col. 3

21 Baby initiates; baby responds
Row 6; Col. 5

22 Baby initiates; baby initiates
Row 6; Col. 6

23 Baby initiates; parent accepts, amplifies
Row 6; Col. 12,13

24 Baby initiates; parent elicits, initiates, directs
Row 6; Col. 14,16,17

25 Parent accepts, amplifies; baby warms, accepts
Rows 12,13; Col. 1,2

26 Parent accepts, amplifies; baby amplifies
Rows 12,13; Col. 3

27 Parent accepts, amplifies; baby responds
Rows 12,13; Col. 5

28 Parent accepts, amplifies; baby initiates
Rows 12,13; Col. 6

29 Parent elicits,initiates, directs; baby warms, accepts
Rows 14,16,17; Col. 1,2

30 Parent elicits, initiates, directs; baby amplifies
Rows 14,16,17; Col. 3

31 Parent elicits, initiates, directs; baby responds
Rows 14,16,17; Col. 5

32 Parent elicits, initiates, directs; baby initiates
Rows 14,16,17; Col. 6

factorial design (see Figure 1). The full model for this design is shown in Table 4. The Expected Mean Squares and the unbiased Estimators of the Components of Variance for each source of variance are shown in Table 5. Both coders and subjects were considered random; Parent Sex was considered fixed. Throughout this study fixed facets are designated by capitalizing the facet name (i.e., Parent Sex); random facets are designated by use of all lower case letters (i.e., coder). The Expected Mean Squares and the Estimates of the Components of Variance were computed using the BMD Analysis of Variance Program BMD08V (Dixon, 1974).

The intercoder agreement for ratings of factor scores of parent-infant interaction for each session was evaluated through the calculation of four generalizability coefficients. The first was $\hat{\rho}^2$ (b, c, P) which provided a measure of intercoder agreement when generalization was intended for all coders, all subjects and the two specific levels of Parent Sex. The equation for computing this coefficient is shown in Table 6.

If this coefficient was low it was necessary to reduce the universe of generalization by looking at a first order coefficient of which there were two. The second coefficient was $\hat{\rho}^2$ (b, c*, P) which provided a measure of generalizability when the universe of generalization was to all subjects and both levels of Parent Sex, but only to the specific coders used in this study. This is indicated by an asterisk for the facet of coder and was adapted from Cronbach et al. (1972). Notice that the numerator (the definition of universe score) changes in that the variance due to coder x block is added to the variance due to block

	$coder_1$		$coder_2$	
	Parent Sex_1	Parent Sex_2	Parent Sex_1	Parent Sex_2
$subject_1$				
$subject_2$				
.				
.				
.				
.				
.				
$subject_n$				

Figure 1

Block Diagram for Randomized Block Factorial Design for Intercoder Agreement Analysis

Table 4

Full Model and Notation for Randomized Block Factorial Design for Intercoder Agreement Analysis

$$X_{ijk} = U + c_i + P_j + b_k + cP_{ij} + cb_{ik} + Pb_{jk} + E_{ijk}$$

where

U = grand mean

c_i = effect of coder i;

P_j = effect of Parent j;

b_k = a constant associated with block k, where a block is an infant;

cP_{ij} = effect of interaction of coder i with Parent j;

cb_{ik} = effect of interaction of coder i with block k;

Pb_{jk} = effect of interaction of Parent j with block k;

E_{ijk} = experimental error;

and there are

Q levels of c_i

R levels of P_j

N levels of b_k

Table 5

Expected Mean Squares and Estimates of Variance Components for Each Source of Variance

Source	Expected Mean Square	Source	Estimate of Variance Components
coder	$RN\hat{\sigma}_{c}^{2} + R\hat{\sigma}_{cb}^{2}$	$\hat{\sigma}_{c}^{2}$	$\frac{1}{RN}(MS_{c} - MS_{cb})$
Parent Sex	$QN\hat{\sigma}_{P}^{2} + N\hat{\sigma}_{cP}^{2} + Q\hat{\sigma}_{Pb}^{2} + \hat{\sigma}_{res}^{2}$	$\hat{\sigma}_{P}^{2}$	$\frac{1}{QN}(MS_{P}-MS_{cP}-MS_{Pb}-MS_{res})$
block	$QR\hat{\sigma}_{b}^{2} + R\hat{\sigma}_{cb}^{2}$	$\hat{\sigma}_{b}^{2}$	$\frac{1}{QR}(MS_{b} - MS_{cb})$
cP	$N\hat{\sigma}_{cP}^{2} + \hat{\sigma}_{res}^{2}$	$\hat{\sigma}_{cP}^{2}$	$\frac{1}{N}(MS_{cP} - MS_{res})$
cb	$R\hat{\sigma}_{cb}^{2}$	$\hat{\sigma}_{cb}^{2}$	$\frac{1}{R}(MS_{cb})$
Pb	$Q\hat{\sigma}_{Pb}^{2} + \hat{\sigma}_{res}^{2}$	$\hat{\sigma}_{Pb}^{2}$	$\frac{1}{Q}(MS_{Pb} - MS_{res})$
residual	$\hat{\sigma}_{res}^{2}$	$\hat{\sigma}_{res}^{2}$	(MS_{res})

Table 6

Description and Algorithm for Estimation of Generalizability Coefficients for Intercoder Agreement Analysis[a]

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizablity Coefficient[b]
$\hat{\rho}^2(b,c,P)$	Generalization intended over all subjects and coders, and to both levels of Parent Sex	$\frac{\hat{\sigma}_b^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c^*,P)$	Generalization intended over all subjects and both levels of Parent Sex, but to only the specific coders in this study	$\frac{\hat{\sigma}_b^2 + \frac{1}{Q}\sigma_{cb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c,P^*)$	Generalization intended over all subjects and coders, but to only one level of Parent Sex	$\frac{\hat{\sigma}_b^2 + \frac{1}{R}\sigma_{Pb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c^*,P^*)$	Generalization intended over all subjects, but to the specific set of coders in this study and to one level of Parent Sex	$\frac{\hat{\sigma}_b^2 + \hat{\sigma}_{cb}^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2}{\hat{\sigma}_x^2}$

a See Table 4 for notation

b $\hat{\sigma}_x^2 = \hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2 + \frac{1}{QR}\hat{\sigma}_{residual}^2$

but the denominator (the definition of observed score) does not change. The generalizability coefficient will increase to the extent that ratings of subjects change when observed by either of the coders. However, this coefficient will represent a smaller universe in that the facet of coder must be considered fixed when estimating the generalizability of the factor scores. Therefore, it was only considered if the generalizability coefficient $\hat{\rho}^2(b,c,P)$ was less than .70. This equation is also shown in Table 6.

The third was $\hat{\rho}^2(b,c,P^*)$, which provided a measure of generalizability when the universe of generalization was to all subjects and coders, but to only one level of Parent Sex. This equation is also shown in Table 6.

If one of these latter two coefficients was still not acceptable, a fourth coefficient, $\hat{\rho}^2(b,c^*,P^*)$, could be considered. The universe of generalization in this case was to all subjects, this specific set of coders and to a specific level of Parent Sex. Again, the equation is shown in Table 6.

Reliability

Subjects crossed with coders. The second analysis was of the reliability of factor scores of parent-infant interaction using matched data (i.e., all subjects were observed by both coders). This is the type of reliability study recommended by Medley and Mitzel (1963) and Cronbach et al. (1963). Six families were observed by both coders for each of three sessions (4 boys, 2 girls).

The data were analyzed using a randomized block factorial design (see Figure 2) with the facets being coder, Parent Sex and Occasion.

	$coder_1$						$coder_2$					
	Parent Sex_1			Parent Sex_2			Parent Sex_1			Parent Sex_2		
	Occ_1	Occ_2	Occ_3	Occ_1	Occ_2	Occ_3	Occ_1	Occ_2	Occ_3	Occ_1	Occ_2	Occ_3
$subject_1$												
$subject_2$												
.												
.												
.												
.												
.												
$subject_n$												

Figure 2

Block Diagram for Randomized Block Factorial Design
for Reliability Analysis
When Subjects Were Crossed With Coders

This analysis allows the direct comparison of the relative effects of each of these facets through the comparison of variance explained (i.e., percent of total sums of squares) or through comparison of the estimates of variance components as well as the comparison of the respective generalizability coefficients. However, all three of these comparisons are fairly crude and do not allow for significance testing. The full model for this design is shown in Table 7. The Expected Mean Squares and the unbiased Estimates of the Components of Variance are shown in Table 8. Coders and subjects were considered random; Parent Sex and Occasion were considered fixed. Again, the Expected Mean Squares and the Estimates of the Components of Variance were computed using the BMD Analysis of Variance Program BMD08V (Dixon, 1974).

In this analysis, which was the method for analyzing Objective 2, eight different generalizability coefficients were determined for each factor. The first coefficient was $\hat{\rho}^2(s,c,P,O)$ which provided an estimate of the reliability of ratings of subjects on factor scores of parent-infant interaction when generalization was intended for all coders and subjects, two specific levels of Parent Sex and three specific Occasions. The equation for computing this coefficient is shown in Table 9.

If this coefficient was low it was necessary to look to a first-order coefficient of which there were three. The second was $\hat{\rho}^2(b,c^*,P,O)$ which provided an estimate of the reliability of ratings of subjects on factor scores of parent-infant interaction when generalization was intended over all subjects and to all levels of Parent Sex and Occasion included in this analysis but to only the specific coders

Table 7

Full Model and Notation for Randomized Block Factorial Design

$$X_{ijkm} = U + c_i + P_j + O_k + b_m + cP_{ij} + cO_{ik} + cb_{im} + PO_{jk} + Pb_{jm} + Ob_{km} + cPO_{ijk} + cPb_{ijm} + cOb_{ikm} + POb_{jkm} + E_{ijkm}$$

where

U = grand mean

c = coder

P = Parent Sex

O = Occasion

b = block, where block equals infant

E = experimental error

and there are

Q levels of c_i

R levels of P_j

T levels of O_k

N levels of b_m

Table 8

Expected Mean Squares and Estimates of Variance Components for Each Source of Variance

Source	Expected Mean Square	Source	Estimate of Variance Component
coder	$RTN\hat{\sigma}_c^2 + RT\hat{\sigma}_{cb}^2$	$\hat{\sigma}_c^2$	$\frac{1}{RTN}(MS_c - MS_{cb})$
Parent Sex	$QTN\hat{\sigma}_P^2 + TN\hat{\sigma}_{cP}^2 + QT\hat{\sigma}_{Pb}^2 + T\hat{\sigma}_{cPb}^2$	$\hat{\sigma}_P^2$	$\frac{1}{QTN}(MS_P - MS_{cP} - MS_{Pb} + MS_{cPb})$
Occasion	$QRN\hat{\sigma}_O^2 + RN\hat{\sigma}_{cO}^2 + QR\hat{\sigma}_{Ob}^2 + R\hat{\sigma}_{cOb}^2$	$\hat{\sigma}_O^2$	$\frac{1}{QRN}(MS_O - MS_{cO} - MS_{Ob} + MS_{cOb})$
block	$QRT\hat{\sigma}_b^2 + RT\hat{\sigma}_{cb}^2$	$\hat{\sigma}_b^2$	$\frac{1}{QRT}(MS_b - MS_{cb})$
cP	$TN\hat{\sigma}_{cP}^2 + T\hat{\sigma}_{cPb}^2$	$\hat{\sigma}_{cP}^2$	$\frac{1}{TN}(MS_{cP} - MS_{cPb})$
cO	$RN\hat{\sigma}_{cO}^2 + R\hat{\sigma}_{cOb}^2$	$\hat{\sigma}_{cO}^2$	$\frac{1}{RN}(MS_{cO} - MS_{cOb})$
cb	$RT\hat{\sigma}_{cb}^2$	$\hat{\sigma}_{cb}^2$	$\frac{1}{RT}(MS_{cb})$
PO	$QN\hat{\sigma}_{PO}^2 + N\hat{\sigma}_{cPO}^2 + Q\hat{\sigma}_{POb}^2 + \hat{\sigma}_{residual}^2$	$\hat{\sigma}_{PO}^2$	$\frac{1}{QN}(MS_{PO} - MS_{cPO} - MS_{POb} + MS_{residual})$
Pb	$QT\hat{\sigma}_{Pb}^2 + T\hat{\sigma}_{cPb}^2$	$\hat{\sigma}_{Pb}^2$	$\frac{1}{QT}(MS_{Pb} - MS_{cPb})$
Ob	$QR\hat{\sigma}_{Ob}^2 + R\hat{\sigma}_{cOb}^2$	$\hat{\sigma}_{Ob}^2$	$\frac{1}{QR}(MS_{Ob} - MS_{cOb})$
cPO	$N\hat{\sigma}_{cPO}^2 + \hat{\sigma}_{residual}^2$	$\hat{\sigma}_{cPO}^2$	$\frac{1}{N}(MS_{cPO} - MS_{residual})$
cPb	$T\hat{\sigma}_{cPb}^2$	$\hat{\sigma}_{cPb}^2$	$\frac{1}{T}(MS_{cPb})$
cOb	$R\hat{\sigma}_{cOb}^2$	$\hat{\sigma}_{cOb}^2$	$\frac{1}{R}(MS_{cOb})$
POb	$Q\hat{\sigma}_{POb}^2 + \hat{\sigma}_{residual}^2$	$\hat{\sigma}_{POb}^2$	$\frac{1}{Q}(MS_{POb} - MS_{residual})$
residual	$\hat{\sigma}_{residual}^2$	$\hat{\sigma}_{res}^2$	$(MS_{residual})$

Table 9

Description and Algorithms for Estimation of Generalizability Coefficients When Subjects Are Crossed With Coders[a]

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficient[b]
$\hat{\rho}^2(b,c,P,O)$	Generalization intended over all subjects and coders and to all levels of Parent Sex and Occasion included in this analysis	$\frac{\hat{\sigma}_b^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c^*,P,O)$	Generalization intended over all subjects and to all levels of Parent Sex and Occasion included in this analysis, but to only the specific coders used in this study	$\frac{\hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c,P^*,O)$	Generalization intended only to one level of Parent Sex, but to all levels of subjects and coders and all levels of Occasion used in this analysis	$\frac{\hat{\sigma}_b^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c,P,O^*)$	Generalization intended only to one level of Occasion, but over all subjects and coders and both levels of Parent Sex	$\frac{\hat{\sigma}_b^2 + \frac{1}{T}\hat{\sigma}_{Pb}^2}{\hat{\sigma}_x^2}$

Table 9 - continued

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficient[b]
$\hat{\rho}^2(b,c^*,P^*,0)$	Generalization intended to specific coders used in this study, to one level of Parent Sex, but to all subjects and all levels of Occasion included in the design	$\dfrac{\hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2 + \frac{1}{QR}\hat{\sigma}_{cPb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,C^*,P,0^*)$	Generalization intended to specific coders used in this study, to one level of Occasion, but to all subjects and all level of Occasion included in the design	$\dfrac{\hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2 + \frac{1}{T}\hat{\sigma}_{Ob}^2 + \frac{1}{QT}\hat{\sigma}_{cOb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c,P^*,0^*)$	Generalization intended only to one level of Parent Sex and one level of Occasion, but to all subjects and coders	$\dfrac{\hat{\sigma}_b^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2 + \frac{1}{T}\hat{\sigma}_{Pb}^2 + \frac{1}{RT}\hat{\sigma}_{POb}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(b,c^*,P^*,0^*)$	Generalization intended to all subjects but to specific coders used in this study, to one level of Parent Sex and one level of Occasion	$\dfrac{\hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2 + \frac{1}{T}\hat{\sigma}_{Pb}^2 + \frac{1}{QR}\hat{\sigma}_{cPb}^2 + QT\,\hat{\sigma}_{cOb}^2 + RT\,\hat{\sigma}_{POb}^2}{\hat{\sigma}_x^2}$

a See explanation of notation in Table 6

b $\hat{\sigma}_x^2 = \hat{\sigma}_b^2 + \frac{1}{Q}\hat{\sigma}_{cb}^2 + \frac{1}{R}\hat{\sigma}_{Pb}^2 + \frac{1}{T}\hat{\sigma}_{Ob}^2 + \frac{1}{QR}\hat{\sigma}_{cPb}^2 + \frac{1}{QT}\hat{\sigma}_{cOb}^2 + \frac{1}{RT}\hat{\sigma}_{POb}^2 + \frac{1}{QRT}\hat{\sigma}_{residual}^2$

used in this study. The third coefficient was $\hat{\rho}^2(b,c,P^*,O)$ which provided an estimate of the reliability of ratings of subjects on factor scores of parent-infant interaction when generalization was intended to only one level of Parent Sex, but everything else was as in the first coefficient. The fourth coefficient was $\hat{\rho}^2(b,c,P,O^*)$ and provided an estimate of reliability when generalization was intended to only one level of Occasion, but everything else was as in the first coefficient. The equations for computing these coefficients are shown in Table 9.

If one of these three coefficients was not adequate it was necessary to look to a second-order coefficient, thereby further reducing the universe of generalization. There were again three coefficients. The fifth coefficient was $\hat{\rho}^2(b,c^*,P^*,O)$, which provided an estimate of the reliability of factor scores when generalization was intended to the specific coders used in this study and one level of Parent Sex, but to all subjects and all levels of Occasion used in this analysis. The fifth was $\hat{\rho}^2(b,c^*,P,O^*)$, which provided an estimate of reliability when generalizability was intended to the specific coders used in this study and one level of Occasion, but to all subjects and both levels of Parent Sex. The seventh coefficient was $\hat{\rho}^2(b,c,P^*,O^*)$, which provided an estimate of reliability when generalization was intended to only one level of Parent Sex and one level of Occasion, but to all subjects and all coders. The equations for these three coefficients are also shown in Table 9.

If one of these coefficients was still not adequate, it was necessary to look to the third-order coefficient $\hat{\rho}^2(b,c^*,P^*,O^*)$, which provided an estimate of the reliability of factor scores when generalization

was intended to all subjects, but to the specific coders used in this study, one level of Parent Sex and one level of Occasion. The equation for this coefficient is shown in Table 9.

Subjects nested within coders. The factor scores were also analyzed separately using a five-facet split-plot design with each facet having two levels (see Figure 3) and was the method of analysis for Objective 3. The analysis of the generalizability of factors of parent-infant interaction using this complex design has the distinct advantage of investigating the reliability of these factors in a sophisticated manner without the additional expense of double-coding (i.e., each family coded twice, once by each coder) all families for all sessions. However, this design does have a serious flaw in that variance due to coder is completely confounded with the variance due to the group to which the coder was assigned (Kirk, 1968). If there were significant differences between groups, then these differences could mask or exaggerate differences between coders.

Campbell and Stanley (1963) have suggested that different designs (each with different flaws) could be patched together such that the overall analysis is less flawed than any one design used by itself. In this specific case this can be done by using information from the two previous analyses where some portion of the families were double coded at each session in order to get some idea of coder effects before the more complex design is used.

The 28 families were first rank-ordered according to the average SES Index for the family calculated by adding the father's Index to the mother's Index and dividing by two. Taking the families two at

			PARENT SEX_1 (MOTHER)				PARENT SEX_2 (FATHER)			
			AGE_1 (19 & 25 WEEKS)		AGE_2 (37 & 43 WEEKS)		AGE_1 (19 & 25 WEEKS)		AGE_2 (47 & 43 WEEKS)	
			$TASK_1$ (LOW MASTERY)	$TASK_2$ (HIGH MASTERY)	$TASK_1$ (LOW MASTERY)	$TASK_2$ (HIGH MASTERY)	$TASK_1$ (LOW MASTERY)	$TASK_2$ (HIGH MASTERY)	$TASK_1$ (LOW MASTERY)	$TASK_2$ (HIGH MASTERY)
$CODER_2$	MALE INFANT	S_1 . . S_7								
	FEMALE INFANT	S_8 . . S_{14}								
$CODER_1$	MALE INFANT	S_{15} . . S_{21}								
	FEMALE INFANT	S_{22} . . S_{28}								

Figure 3

Block Diagram for Five-facet Split-plot Design

a time, beginning with the highest Index (i.e., lowest score), the families were randomly assigned to one of two groups, balancing for sex of infant. One observer was then randomly assigned to code all of the families in one group. The Bayley Mental Development Index (Bayley, 1969) was administered to the infants at age 12 months and a t-test was done to determine if there were significant differences on this scale between the infants in the two groups.

The facets used in the split-plot design were coder, Sex of Parent, Sex of Infant, Age of Infant and Type of Task. The levels for Age of Infant were determined by separating the sessions into two groups: (1) less than seven months (19 and 25 weeks) and (2) over seven months (37 and 43 weeks).

Levels for Type of Task were established through analysis of a measure (Task Mastery) from the Home Scale Observation Instrument (Watts and Barnett, 1971), which was a measure of the child's ability to correctly perform the task. An analysis of the means for each session indicated that two tasks were more readily accomplished by the infants ($\bar{x}_{19} = 4.17$; $\bar{x}_{37} = 2.94$) and two tasks were performed less well ($\bar{x}_{25} = .67$; $\bar{x}_{43} = .28$). Therefore, the Type of Task was divided into two levels, one of high mastery (19 and 37 weeks) and a second of low mastery (25 and 43 weeks). The full model for this design is shown in Table 10. The Expected Mean Squares and the unbiased Estimators of the Components of Variance for each source of variance are shown in Table 11. All variables were considered fixed except coders and subjects, which were considered random. Again, fixed facets were designated by capitalizing the facet (i.e., Parent Sex); random

Table 10

Full Model and Notation for
Five-facet Split-plot Design

$$X_{hjklo} = U + c_h + I_k + cI_{hk} + sub/cI_{m(hk)} + P_j + cP_{hj} + PI_{jk} + cPI_{hjk}$$
$$+ Ps/cI_{jm(hk)} + A_l + cA_{hl} + IA_{kl} + cIA_{hkl} + As/cI_{lm(hk)} + T_o$$
$$+ cT_{ho} + IT_{ko} + cIT_{hko} + Ts/cI_{om(hk)} + PA_{jl} + cPA_{hjl} + PIA_{jkl}$$
$$+ cPIA_{hjkl} + PAs/cI_{jlm(hk)} + PT_{jl} + cPT_{hjo} + PIT_{jko} + cPIT_{hjko}$$
$$+ PTs/cI_{jom(hk)} + AT_{lo} + cAT_{hlo} + IAT_{klo} + cIAT_{hklo} + ATs/cI_{lom(hk)}$$
$$+ PAT_{jlo} + cPAT_{hjlo} + PIAT_{jklo} + cPIAT_{hjklo} + PATs/cI_{jlom(hk)}$$
$$+ E_{g(hjklo)}$$

where

U = Grand mean

c = coder

I = Infant Sex

P = Parent Sex

A = Infant Age Grouping

T = Type of Task

s = subject

E = Experimental Error

and there are

V levels of c_h

U levels of I_k

R levels of P_j

O levels of A_l

W levels of T_o

N subjects

Table 11

Expected Mean Squares and Estimate of Variance Components for Each Source of Variance

Source	Expected Mean Square	Source	Estimate of Variance Components
c	$NVURQ\sigma^2_c + VUQ\sigma^2_{s(cI)}$	$\hat{\sigma}^2_c$	$(\frac{1}{NVURQ})(MS_c - MS_{s(cI)})$
I	$NVURW\sigma^2_I + NVUQ\sigma^2_{cI} + VUQ\sigma^2_{s(cI)}$	$\hat{\sigma}^2_I$	$(\frac{1}{NVURW})(MS_I - MS_{cI})$
cI	$NVUQ\sigma^2_{cI} + VUQ\sigma^2_{s(cI)}$	$\hat{\sigma}^2_{cI}$	$(\frac{1}{NVUQ})(MS_{cI} - MS_{s(cI)})$
sub/cI	$VUQ\sigma^2_{s(cI)}$	$\hat{\sigma}^2_{s(cI)}$	$(\frac{1}{VUQ})(MS_{s(cI)})$
P	$NVURW\sigma^2_P + NVUR\sigma^2_{cP} + VU\sigma^2_{Ps(cI)}$	$\hat{\sigma}^2_P$	$(\frac{1}{NVURW})(MS_P - MS_{cP})$
cP	$NVUR\sigma^2_{cP} + VU\sigma^2_{Ps(cI)}$	$\hat{\sigma}^2_{cP}$	$(\frac{1}{NVUR})(MS_{cP} - MS_{Ps(cI)})$
PI	$NVUW\sigma^2_{PI} + NVU\sigma^2_{cPI} + VU\sigma^2_{Ps(cI)}$	$\hat{\sigma}^2_{PI}$	$(\frac{1}{NVUW})(MS_{PI} - MS_{cPI})$
cPI	$NVU\sigma^2_{cPI} + VU\sigma^2_{Ps(cI)}$	$\hat{\sigma}^2_{cPI}$	$(\frac{1}{NVU})(MS_{cPI} - MS_{Ps(cI)})$
P_*sub/cI	$VU\sigma^2_{Ps(cI)}$	$\hat{\sigma}^2_{Ps(cI)}$	$(\frac{1}{VU})(MS_{Ps(cI)})$
A	$NVRQW\sigma^2_A + NVRQ\sigma^2_{cA} + VQ\sigma^2_{As(cI)}$	$\hat{\sigma}^2_A$	$(\frac{1}{NVRQW})(MS_A - MS_{cA})$
cA	$NVRQ\sigma^2_{cA} + VQ\sigma^2_{As(cI)}$	$\hat{\sigma}^2_{cA}$	$(\frac{1}{NVRQ})(MS_{cA} - MS_{As(cI)})$
IA	$NVQW\sigma^2_{IA} + NVQ\sigma^2_{cIA} + VQ\sigma^2_{As(cI)}$	$\hat{\sigma}^2_{IA}$	$(\frac{1}{NVQW})(MS_{IA} - MS_{cIA})$
cIA	$NVQ\sigma^2_{cIA} + VQ\sigma^2_{As(cI)}$	$\hat{\sigma}^2_{cIA}$	$(\frac{1}{NVQ})(MS_{cIA} - MS_{As(cI)})$
A_*sub/cI	$VQ\sigma^2_{As(cI)}$	$\hat{\sigma}^2_{As(cI)}$	$(\frac{1}{VQ})(MS_{As(cI)})$
T	$NURQW\sigma^2_T + NURP\sigma^2_{cT} + UQ\sigma^2_{Ts(cI)}$	$\hat{\sigma}^2_T$	$(\frac{1}{NURQW})(MS_T - MS_{cT})$
cT	$NURQ\sigma^2_{cT} + UQ\sigma^2_{Ts(cI)}$	$\hat{\sigma}^2_{cT}$	$(\frac{1}{NURW})(MS_{cT} - MS_{Ts(cI)})$
IT	$NUQW\sigma^2_{IT} + NUQ\sigma^2_{cIT} + UQ\sigma^2_{Ts(cI)}$	$\hat{\sigma}^2_{IT}$	$(\frac{1}{NUQW})(MS_{IT} - MS_{cIT})$
cIT	$NUQ\sigma^2_{cIT} + UQ\sigma^2_{Ts(cI)}$	$\hat{\sigma}^2_{cIT}$	$(\frac{1}{NUQ})(MS_{cIT} - MS_{Ts(cI)})$
T_*sub/cI	$UQ\sigma^2_{Ts(cI)}$	$\hat{\sigma}^2_{Ts(cI)}$	$(\frac{1}{UQ})(MS_{Ts(cI)})$
PA	$NVRW\sigma^2_{PA} + NVR\sigma^2_{cPA} + V\sigma^2_{PAs(cI)}$	$\hat{\sigma}^2_{PA}$	$(\frac{1}{NVRW})(MS_{PA} - MS_{cPA})$
cPA	$NVR\sigma^2_{cPA} + V\sigma^2_{PAs(cI)}$	$\hat{\sigma}^2_{cPA}$	$(\frac{1}{NVR})(MS_{cPA} - MS_{PAs(cI)})$
PIA	$NVW\sigma^2_{PIA} + NV\sigma^2_{cPIA} + V\sigma^2_{PAs(cI)}$	$\hat{\sigma}^2_{PIA}$	$(\frac{1}{NVW})(MS_{PIA} - MS_{cPIA})$

Table 11 - continued

Source	Expected Mean Square	Source	Estimate of Variance Components
cPIA	$NV\sigma^2_{cPIA} + V\sigma^2_{PAs(cI)}$	$\hat{\sigma}^2_{cPIA}$	$(\frac{1}{NV})(MS_{cPIA} - MS_{PAs(cI)})$
PA*sub/cI	$V\sigma^2_{PAs(cI)}$	$\hat{\sigma}^2_{PAs(cI)}$	$(\frac{1}{V})(MS_{PAs(cI)})$
PT	$NURW\sigma^2_{PT} + NUR\sigma^2_{cPT} + U\sigma^2_{PTs(cI)}$	$\hat{\sigma}^2_{PT}$	$(\frac{1}{NURW})(MS_{PT} - MS_{cPT})$
cPT	$NUR\sigma^2_{cPT} + U\sigma^2_{PTs(cI)}$	$\hat{\sigma}^2_{cPT}$	$(\frac{1}{NUR})(MS_{cPT} - MS_{PTs(cI)})$
PIT	$NUW\sigma^2_{PIT} + NU\sigma^2_{cPIT} + U\sigma^2_{PTs(cI)}$	$\hat{\sigma}^2_{PIT}$	$(\frac{1}{NUW})(MS_{PIT} - MS_{cPIT})$
cPIT	$NU\sigma^2_{cPIT} + U\sigma^2_{PTs(cI)}$	$\hat{\sigma}^2_{cPIT}$	$(\frac{1}{NU})(MS_{cPIT} - MS_{PTs(cI)})$
PT*sub/cI	$U\sigma^2_{PTs(cI)}$	$\hat{\sigma}^2_{PTs(cI)}$	$(\frac{1}{U})(MS_{PTs(cI)})$
AT	$NRQW\sigma^2_{AT} + NRQ\sigma^2_{cAT} + Q\sigma^2_{ATs(cI)}$	$\hat{\sigma}^2_{AT}$	$(\frac{1}{NRQW})(MS_{AT} - MS_{cAT})$
cAT	$NRQ\sigma^2_{cAT} + Q\sigma^2_{ATs(cI)}$	$\hat{\sigma}^2_{cAT}$	$(\frac{1}{NRQ})(MS_{cAT} - MS_{ATs(cI)})$
IAT	$NQW\sigma^2_{IAT} + NQ\sigma^2_{cIAT} + Q\sigma^2_{ATs(cI)}$	$\hat{\sigma}^2_{IAT}$	$(\frac{1}{NQW})(MS_{IAT} - MS_{cIAT})$
cIAT	$NQ\sigma^2_{cIAT} + Q\sigma^2_{ATs(cI)}$	$\hat{\sigma}^2_{cIAT}$	$(\frac{1}{NQ})(MS_{cIAT} - MS_{ATs(cI)})$
AT*sub/cI	$Q\sigma^2_{ATs(cI)}$	$\hat{\sigma}^2_{ATs(cI)}$	$(\frac{1}{Q})(MS_{ATs(cI)})$
PAT	$NRW\sigma^2_{PAT} + NR\sigma^2_{cPAT} + \sigma^2_{PATs(cI)}$	$\hat{\sigma}^2_{PAT}$	$(\frac{1}{NRW})(MS_{PAT} - MS_{cPAT})$
cPAT	$NR\sigma^2_{cPAT} + \sigma^2_{PATs(cI)}$	$\hat{\sigma}^2_{cPAT}$	$(\frac{1}{NR})(MS_{cPAT} - MS_{PATs(cI)})$
PIAT	$NW\sigma^2_{PIAT} + N\sigma^2_{cPIAT} + \sigma^2_{PATs(cI)}$	$\hat{\sigma}^2_{PIAT}$	$(\frac{1}{NW})(MS_{PIAT} - MS_{cPIAT})$
cPIAT	$N\sigma^2_{cPIAT} + \sigma^2_{PATs(cI)}$	$\hat{\sigma}^2_{cPIAT}$	$(\frac{1}{N})(MS_{cPIAT} - MS_{PATs(cI)})$
AT*sub/cI	$\sigma^2_{PATs(cI)}$	$\sigma^2_{PATs(cI)}$	$(MS_{PATs(cI)})$

facets were designated through the use of all lower case letters (i.e., coder).

In this analysis eight different generalizability coefficients were determined for each factor. The first coefficient was ρ^2(s,c,I, P,A,T) which provided an estimate of the reliability of ratings of subjects on factor scores of parent-infant interaction when generalization was intended for all coders, both levels of Infant Sex, both levels of Parent Sex, the two Age Groupings of infants considered in this study and the two levels of Tasks considered. The equation for computing the coefficient is shown in Table 12.

The second coefficient was ρ^2(s,c,I,P*,A,T) which provided an estimate of the reliability of ratings of subjects on factor scores of parent-infant interaction when generalization was intended to only one level of Parent Sex, but everything else remained the same. The equation for computing this coefficient is also shown in Table 12.

The third coefficient was ρ^2(s,c,I,P, A*,T) which provided an estimate of ratings of subjects when generalization was intended to only one level of infant Age Grouping, but everything was as in the first coefficient. Again, the equation for computing this coefficient is shown in Table 12.

The fourth coefficient was ρ^2(s,c,I,P,A,T*) which provided an estimate of the reliability of ratings of subjects when generalization was intended to only one level of Task, but again, everything was as in the first equation. The equation for computing this coefficient is also shown in Table 12.

If the first generalizability coefficient was less than .70, it

Table 12

Description and Algorithms for Estimation of Generalizability Coefficients

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficient[b]
$\hat{\rho}^2(s,c,I,P,A,T)$	Generalization intended over all facets	$\dfrac{\hat{\sigma}_{sub(cI)}^{\;2}}{\hat{\sigma}_X^{\;2}}$
$\hat{\rho}^2(s,c,I,P^*,A,T)$	Generalization not intended over facet of Parent Sex	$\dfrac{\hat{\sigma}_{sub(cI)}^{\;2} + \frac{1}{R}\hat{\sigma}_{P*sub(cI)}^{\;2}}{\hat{\sigma}_X^{\;2}}$
$\hat{\rho}^2(s,c,I,P,A^*,T)$	Generalization not intended over facet of Infant Age Grouping	$\dfrac{\hat{\sigma}_{sub(cI)}^{\;2} + \frac{1}{O}\hat{\sigma}_{A*sub(cI)}^{\;2}}{\hat{\sigma}_X^{\;2}}$
$\hat{\rho}^2(s,c,I,P,A,T^*)$	Generalization not intended over facet of Type of Task	$\dfrac{\hat{\sigma}_{sub(cI)}^{\;2} + \frac{1}{W}\hat{\sigma}_{T*sub(cI)}^{\;2}}{\hat{\sigma}_X^{\;2}}$

Table 12 - continued

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficients[b]
$\hat{\rho}^2(s,c,I,P^*,A^*,T)$	Generalization not intended over facets of Parent Sex or Infant Age Grouping	$\dfrac{\hat{\sigma}_{sub(cI)}^2 + \frac{1}{R}\hat{\sigma}_{P*sub(cI)}^2 + \frac{1}{O}\hat{\sigma}_{A*sub(cI)}^2 + \frac{1}{RO}\hat{\sigma}_{PA*sub(cI)}^2}{\hat{\sigma}_X^2}$
$\hat{\rho}^2(s,c,I,P^*,A,T^*)$	Generalization not intended over facets of Parent Sex or Type of Task	$\dfrac{\hat{\sigma}_{sub(cI)}^2 + \frac{1}{R}\hat{\sigma}_{P*sub(cI)}^2 + \frac{1}{W}\hat{\sigma}_{T*sub(cI)}^2 + \frac{1}{RW}\hat{\sigma}_{PT*sub(cI)}^2}{\hat{\sigma}_X^2}$
$\hat{\rho}^2(s,c,I,P,A^*,T^*)$	Generalization not intended over facets of Infant Age Grouping or Type of Task	$\dfrac{\hat{\sigma}_{sub(cI)}^2 + \frac{1}{O}\hat{\sigma}_{A*sub(cI)}^2 + \frac{1}{W}\hat{\sigma}_{T*sub(cI)}^2 + \frac{1}{OW}\hat{\sigma}_{AT*sub(cI)}^2}{\hat{\sigma}_X^2}$

Table 12 - continued

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficients[b]
$\hat{\rho}^2(s,c,I,P^*,A^*,T^*)$	Generalizability not intended over facets of Parent Sex, Infant Age Grouping, or Type of Task	$\dfrac{\hat{\sigma}_{sub(c)}^2 + \frac{1}{R}\hat{\sigma}_{P*sub(cI)}^2 + \frac{1}{O}\hat{\sigma}_{A*sub(cI)}^2 + \frac{1}{W}\hat{\sigma}_{T*sub(cI)}^2 + \frac{1}{RO}\hat{\sigma}_{PA*sub(cI)}^2 + \frac{1}{RW}\hat{\sigma}_{PT*sub(cI)}^2 + \frac{1}{OW}\hat{\sigma}_{AT*sub(cI)}^2}{\hat{\sigma}_X^2}$

a See explanation of notation in Table 6

b $\hat{\sigma}_X^2 = \hat{\sigma}_{sub(cI)}^2 + \frac{1}{R}\hat{\sigma}_{P*sub(cI)}^2 + \frac{1}{O}\hat{\sigma}_{A*sub(cI)}^2 + \frac{1}{W}\hat{\sigma}_{T*sub(cI)}^2 + \frac{1}{RO}\hat{\sigma}_{PA*sub(cI)}^2 + \frac{1}{RW}\hat{\sigma}_{PT*sub(cI)}^2$

$+ \frac{1}{OW}\hat{\sigma}_{AT*sub(cI)}^2 + \frac{1}{ROW}\hat{\sigma}_{PAT*sub(cI)}^2 + \frac{1}{U}\hat{\sigma}_{I}^2 + \frac{1}{UR}\hat{\sigma}_{IP}^2 + \frac{1}{UO}\hat{\sigma}_{IA}^2 + \frac{1}{UW}\hat{\sigma}_{IT}^2 + \frac{1}{URO}\hat{\sigma}_{IPA}^2$

$+ \frac{1}{URW}\hat{\sigma}_{IPT}^2 + \frac{1}{UOW}\hat{\sigma}_{IAT}^2 + \frac{1}{UROW}\hat{\sigma}_{IPAT}^2 + \frac{1}{V}\hat{\sigma}_{c}^2 + \frac{1}{VU}\hat{\sigma}_{cI}^2 + \frac{1}{VR}\hat{\sigma}_{cP}^2 + \frac{1}{VO}\hat{\sigma}_{cA}^2 + \frac{1}{VW}\hat{\sigma}_{cT}^2$

$+ \frac{1}{VUR}\hat{\sigma}_{cIP}^2 + \frac{1}{VUO}\hat{\sigma}_{cIA}^2 + \frac{1}{VUW}\hat{\sigma}_{cIT}^2 + \frac{1}{VRO}\hat{\sigma}_{cPA}^2 + \frac{1}{VRW}\hat{\sigma}_{cPT}^2 + \frac{1}{VOW}\hat{\sigma}_{cAT}^2 + \frac{1}{VRUO}\hat{\sigma}_{cPIA}^2$

$+ \frac{1}{VRUW}\hat{\sigma}_{cPIT}^2 + \frac{1}{VUOW}\hat{\sigma}_{cIAT}^2 + \frac{1}{VROW}\hat{\sigma}_{cPAT}^2 + \frac{1}{VRUOW}\hat{\sigma}_{cPIAT}^2$

was necessary to look at the next three coefficients, each of which represented a first-order interaction with ratings of subjects. If each of these coefficients was less than .70 it meant that one could not generalize subjects' scores when the level of only one facet was held constant and, therefore, one must look to a second-order coefficient. These are shown in Table 12.

Again, it must be emphasized that the generalizability coefficient is defined as the ratio of a universe score variance to observed score variance; the generalizability coefficient should become larger as more of the variance was considered true score variance rather than error variance. However, the universe to which that score would generalize becomes increasingly smaller. If the third-order coefficient $\hat{\rho}^2$(s,c,I, P*,A*,T*) for a factor was not above .70, there were four alternative explanations:

1) the coder facet and/or the infant sex facet was having some systematic influence on the scores;
2) these factor scores varied nonsystematically;
3) these factor scores varied systematically on some facets not considered in this study;
4) some combination of alternatives 1-3.

Summary

Twenty-eight white, first-born infants were video-taped interacting with their mothers and fathers on four separate occasions. Two observers coded the parent-infant interaction using the Reciprocal Category System (RCS) from which 32 measures were derived which accounted for 84% of the total tallies. These measures were factor analyzed and five factor scores were produced.

The intercoder agreement on these factor scores for three of the four occasions was analyzed using a randomized block factorial design using the facets of coder and Parent Sex. The reliability of these factor scores was analyzed using two additional designs. The first of these was also a randomized block factorial design using the facets of coder, Parent Sex and Occasion. The second was a split-plot design where the 28 families were divided into two separate groups of 14 and one observer was assigned to each group. The facets were coder, Infant Sex, Parent Sex, Age of Infant and Type of Task.

CHAPTER IV
ANALYSIS

The purpose of this study was to investigate the reliability of parent-infant interaction through the use of generalizability theory. Video-tapes of parent-infant interaction were coded by two observers using the Reciprocal Category System (RCS). Factor scores were produced and analyzed using three designs. The first design was used to investigate intercoder agreement; the second and third designs were used to investigate reliability. In one of these designs subjects were crossed with coders; in the other, subjects were nested within coders.

The factor analysis, variance components and generalizability coefficients are reported in this chapter.

Development of Factor Scores

The 32 measures from the RCS were factor analyzed using a principal axis solution. Based upon a scree test, the five factors were rotated to the Varimax criteria and the resulting solution accounted for 75% of the common variance (see Tables 13-17). All of the loadings were positive and above .40. Only one variable (14) loaded on two factors (Factors 1 and 3). One interesting finding was that for the most part it was the babies' behavior which determined the factor structure. The parent behaviors of accepting, amplifying, eliciting, initiating and directing all loaded on the same factor dependent upon the baby behavior in the interaction sequence.

Factor 1 (parent-infant interaction; baby amplifies) was composed of both interaction measures and baby behavior measures which involved

Table 13

RCS Factor 1
Parent-Infant Interaction; Baby Amplifies

Item Number	Description	Loading
12	Baby amplifies; parent elicits, initiates, directs 03; 14,16,17	.88
30	Parent elicits, initiates, directs; baby amplifies 14,16,17; 03	.83
26	Parent accepts, amplifies; baby amplifies 12,13; 03	.68
11	Baby amplifies; parent accepts, amplifies 03; 12,13	.66
14	Baby responds; baby amplifies 05; 03	.64
8	Baby amplifies; baby amplifies 03; 03	.58
7	Baby amplifies; baby warms, accepts 03; 01,02	.51
9	Baby amplifies; baby responds 03; 05	.44
10	Baby amplifies; baby initiates 03; 06	.44
2	Baby warms, accepts; baby amplifies 01,02; 03	.43

Eigen Value = 3.41

Table 14

RCS Factor 2
Parent-Infant Interaction; Baby Responds

Item Number	Description	Loading
17	Baby responds; parent accepts, amplifies 05; 12,13	.87
27	Parent accepts, amplifies; baby responds 12,13; 05	.87
18	Baby responds; parent elicits, initiates, directs 05; 14,16,17	.80
31	Parent elicits, initiates, directs; baby responds 14,16,17; 05	.80

Eigen Value = 3.86

Table 15

RCS Factor 3
General Baby Behavior

Item Number	Description	Loading
3	Baby warms, accepts; baby responds 01,02; 05	.67
16	Baby responds; baby inititates 05; 06	.65
4	Baby warms, accepts; baby initiates 01,02; 06	.59
19	Baby initiates; baby warms, accepts 06; 01,02	.59
13	Baby responds; baby warms, accepts 05; 01,02	.55
21	Baby initiates; baby responds 06; 05	.51
1	Baby warms, accepts; baby warms, accepts 01,02; 01,02	.44
15	Baby responds; baby responds 05; 05	.43
14	Baby responds; baby amplifies 05; 03	.41

Eigen Value = 4.29

Table 16

RCS Factor 4
Parent-Infant Interaction; Baby Warms, Accepts

Item Number	Description	Loading
6	Baby warms, accepts; parent elicits, initiates, directs 01,02; 14,16,17	.86
29	Parent elicits, initiates, directs; baby warms, accepts 14,16,17; 01,02	.84
25	Parent accepts, amplifies; baby warms, accepts 12,13; 01,02	.67
5	Baby warms, accepts; parent accepts, amplifies 01,02; 12,13	.61

Eigen Value = 2.73

Table 17

RCS Factor 5
Parent-Infant Interaction; Baby Initiates

Item Number	Description	Loading
32	Parent elicits, initiates, directs; baby initiates 14,16,17; 06	.78
24	Baby initiates; parent elicits, initiates, directs 06; 14,16,17	.77
28	Parent accepts, amplifies; baby initiates 12,13; 06	.76
23	Baby initiates; parent accepts, amplifies 06; 12,13	.72

Eigen Value = 2.98

the baby amplifying. This was a measure of the extent to which the baby was exhibiting extending behavior after the task was presented by the parent.

Factor 2 (parent-infant interaction; baby responds) was composed only of interaction measures and was an indication of the extent to which the baby was responding while being verbally stimulated or responded to by the parent. From an informal analysis of the videotapes it was apparent that the majority of these responses were on-task behaviors (i.e., the babies were behaving in a manner appropriate to the structured teaching situation).

Factor 3 (general baby behavior) was composed exclusively of measures of baby behavior and was a measure of general level of baby activity. It was also some indication that parent verbal behavior was absent. All of the baby behaviors that were used in this analysis were represented in this factor so that it was not possible to be more specific.

Factor 4 (parent-infant interaction; baby warms, accepts) was also composed only of interaction measures and was an indication of the extent to which the baby was either warming or accepting while the parent was exhibiting verbal behavior. These baby behaviors were generally affective and passive as compared to the more active behaviors of responding or initiating, although the majority were on-task behaviors.

Factor 5 (parent-infant interaction; baby initiates) was the third factor which was composed exclusively of interaction measures. It was an indication of the extent to which the baby was emitting behavior

which had no immediate observable antecedent while the parent was verbally engaging the infant. It was generally active, off-task behavior in that the baby was doing something other than the behavior requested in the structured teaching situation.

Analysis of Intercoder Agreement

The intercoder agreement for each of the five factors was analyzed at three separate occasions (25 weeks, 37 weeks, 43 weeks) using randomized block factorial designs. The total sample of the longitudinal project (n = 38) had a mean Bayley Mental Development Quotient (MDQ) of 118.7 with a standard deviation of 12.0 (standardized sample mean = 100 with standard deviation of 15). The infants used in the 25 weeks analysis had a mean MDQ of 116.6 and a standard deviation of 14.3; at 37 weeks the mean MDQ was 115.2 with a standard deviation of 13.8; at 43 weeks the mean MDQ was 112.6 with a standard deviation of 13.1. The components of variance and generalizability coefficients are shown for each factor in Tables 18-22. Coder and block (block = infant) were considered random; Parent Sex was considered fixed. A summary of the generalizability coefficients is shown in Table 23. These coefficients were not an estimate of reliability because variance due to occasion was not considered. A coefficient above .70 was considered an indication of adequate intercoder agreement.

Estimates of Variance Components

Notice that there are negative estimates for each factor, which are especially large for the Parent Sex x block interaction for Factor 2 at 37 and 43 weeks and for block for Factor 4 at 43 weeks. As observed by Llabre (1978), these are obviously poor estimates since a variance is, by definition, non-negative. It is interesting that the majority

Table 18

Components of Variance for Intercoder Agreement Analysis of Factor 1 (Parent-Infant Interaction; Baby Amplifying)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	126.60	19	1	4.07
	Parent Sex	5.04	1	1	.03
	block	191.51	29	13	.51
	cP	.02	0	1	-.30
	cb	164.90	25	13	6.34
	Pb	109.42	17	13	2.10
	residual	54.75	8	13	4.21
	Total	652.24	99	55	
37 weeks	coder	1.74	0	1	-.26
	Parent Sex	23.12	2	1	.17
	block	723.03	52	17	7.85
	cP	.02	0	1	-.26
	cb	189.43	13	17	5.57
	Pb	371.95	27	17	8.56
	residual	80.91	6	17	4.76
	Total	1390.20	100	71	
43 weeks	coder	24.00	2	1	-.10
	Parent Sex	46.18	3	1	1.40
	block	641.69	48	16	3.19
	cP	1.53	0	1	-.36
	cb	437.61	33	16	13.68
	Pb	71.17	5	16	-1.58
	residual	121.82	9	16	7.61
	Total	1344.00	100	67	

Table 19

Components of Variance for Intercoder Agreement Analysis of Factor 2 (Parent-Infant Interaction; Baby Responding)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	51.11	2	1	1.37
	Parent Sex	64.93	3	1	.23
	block	979.63	43	13	15.63
	cP	55.80	2	1	1.45
	cb	166.99	7	13	6.42
	Pb	496.43	22	13	1.36
	residual	461.14	20	13	35.47
	Total	2276.03	99	55	
37 weeks	coder	22.78	1	1	.44
	Parent Sex	44.02	2	1	.78
	block	1373.48	71	17	18.50
	cP	17.90	1	1	.36
	cb	115.64	6	17	3.40
	Pb	157.55	8	17	-1.05
	residual	193.19	10	17	11.36
	Total	1924.56	99	71	
43 weeks	coder	549.49	29	1	15.57
	Parent Sex	4.50	0	1	.22
	block	813.11	42	16	7.68
	cP	3.43	0	16	-.40
	cb	321.37	17	16	10.04
	Pb	61.44	3	16	-3.19
	residual	163.54	9	16	10.22
	Total	1916.88	100	67	

Table 20

Components of Variance for Intercoder Agreement Analysis of Factor 3 (General Baby Behavior)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	6.72	1	1	-.09
	Parent Sex	1.20	0	1	-.64
	block	163.75	27	13	.81
	cP	.16	0	1	-.16
	cb	121.68	20	13	4.68
	Pb	276.71	46	13	9.44
	residual	31.14	5	13	2.40
	Total	601.36	99	55	
37 weeks	coder	119.09	8	1	3.19
	Parent Sex	19.01	1	1	.22
	block	960.23	65	17	13.05
	cP	.11	0	1	-.19
	cb	72.73	5	17	2.14
	Pb	245.14	17	17	5.42
	residual	60.85	4	17	3.58
	Total	1477.16	100	71	
43 weeks	coder	49.13	3	1	.80
	Parent Sex	49.13	3	1	1.01
	block	686.29	45	16	5.21
	cP	11.86	1	1	.09
	cb	353.15	23	16	11.04
	Pb	210.40	14	16	1.39
	residual	165.80	11	16	10.36
	Total	1525.76	100	67	

Table 21

Components of Variance for Intercoder Agreement Analysis
of Factor 4 (Parent-Infant Interaction; Baby Warming, Accepting)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	13.80	1	1	-.84
	Parent Sex	21.13	1	1	-.02
	block	612.56	32	13	2.43
	cP	14.81	1	1	-.74
	cb	486.37	26	13	18.71
	Pb	416.73	22	13	3.44
	residual	327.34	17	13	25.18
	Total	1892.74	100	55	
37 weeks	coder	199.00	6	1	5.09
	Parent Sex	21.02	0	1	.27
	block	2255.97	71	17	29.21
	cP	.01	0	1	-.37
	cb	269.48	9	17	7.93
	Pb	307.63	10	17	5.71
	residual	113.65	4	17	6.69
	Total	3166.76	100	71	
43 weeks	coder	151.80	7	1	2.50
	Parent Sex	91.32	4	1	2.39
	block	778.59	34	16	-4.53
	cP	8.90	0	1	.21
	cb	1068.81	46	16	33.40
	Pb	105.82	5	16	.63
	residual	85.75	4	16	5.36
	Total	2290.99	100	67	

Table 22

Components of Variance for Intercoder Agreement Analysis of Factor 5 (Parent-Infant Interaction; Baby Initiating)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	13.50	0	1	-.53
	Parent Sex	45.18	1	1	-.58
	block	1624.78	51	13	24.14
	cP	3.35	0	1	-.78
	cb	369.53	12	13	14.21
	Pb	938.97	30	13	28.99
	residual	185.25	6	13	14.25
	Total	3180.56	100	55	
37 weeks	coder	.13	0	1	-.08
	Parent Sex	47.69	2	1	.67
	block	1733.30	72	17	24.70
	cP	1.07	0	1	-.26
	cb	53.73	2	17	1.58
	Pb	481.95	20	17	11.34
	residual	96.46	4	17	5.67
	Total	2414.33	100	71	
43 weeks	coder	192.46	9	1	4.25
	Parent Sex	3.31	0	1	-.14
	block	724.69	35	16	-.70
	cP	4.76	0	1	-.32
	cb	769.59	37	16	24.04
	Pb	213.84	10	16	1.59
	residual	162.86	8	16	10.18
	Total	2071.51	99		

Table 23

Generalizability Coefficients for Intercoder Agreement Analysis[a]

Occasion	Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
25 weeks	$\hat{\rho}^2(b,c,P)$	.09	.56	.10	.12	.49
	$\hat{\rho}^2(b,c^*,P)$	.64	.68	.37	.60	.63
	$\hat{\rho}^2(b,c,P^*)$	.27	.59	.66	.21	.78
	$\hat{\rho}^2(b,c^*,P^*)$	.82	.68	.93	.68	.93
37 weeks	$\hat{\rho}^2(b,c,P)$	.49	.80	.74	.77	.74
	$\hat{\rho}^2(b,c^*,P)$	.66	.88	.80	.88	.78
	$\hat{\rho}^2(b,c,P^*)$	.75	.80	.89	.85	.93
	$\hat{\rho}^2(b,c^*,P^*)$	.93	.88	.95	.96	.96
43 weeks	$\hat{\rho}^2(b,c,P)$	.27	.43	.37	0	0
	$\hat{\rho}^2(b,c^*,P)$	.84	.83	.77	.91	.78
	$\hat{\rho}^2(b,c,P^*)$	.27	.43	.42	.02	.05
	$\hat{\rho}^2(b,c^*,P^*)$	.84	.83	.81	.93	.83

a Substituting 0 for negative estimates of variance components

involve the facet of Parent Sex; in only three instances out of 15 is the estimate of variance for the main effect of Parent Sex above 1.00.

The largest estimated variance component for Factor 1 (Parent-infant interaction with baby amplifying) at 25 weeks was for the coder x block interaction ($\hat{\sigma}^2_{cb} = 6.34$); at 37 weeks it was the Parent Sex x block interaction followed by block ($\hat{\sigma}^2_{Pb} = 8.56$, $\hat{\sigma}^2_{b} = 7.85$); at 43 weeks it was again coder x block ($\hat{\sigma}_{cb} = 13.68$). Notice, however, that at 25 weeks the estimated variance component for coder was one of the largest while at 37 and 43 weeks it was negative. This means that at 25 weeks there were differences in the overall means between the two coders but these were negligible at 37 and 43 weeks. In all three instances, though, the estimated variance component for coder x block was one of the largest. This means that the coders were assigning different rankings to subjects irrespective of their overall ratings for all subjects. The former component was part of the observed score variance, while the latter was not (see Table 6).

At 25 weeks the largest estimated variance component for Factors 3 and 5 was the Parent Sex x block interaction and for Factor 2 and 4 it was the residual ($\hat{\sigma}^2_{Pb} = 9.44$, $\hat{\sigma}^2_{Pb} = 28.99$, σ^2 residual = 35.47, $\hat{\sigma}^2$ residual = 25.18, respectively).

At 37 weeks the largest estimated variance component for Factors 2, 3, 4 and 5 was for block ($\hat{\sigma}^2_{b}$ 18.50, $\hat{\sigma}^2_{b} = 13.05$, $\hat{\sigma}^2_{b} = 29.21$, $\hat{\sigma}_{b} = 24.70$, respectively).

At 43 weeks the largest estimated variance component for Factors 3, 4 and 5 was for the coder x block interaction ($\hat{\sigma}^2_{cb} = 11.04$, $\hat{\sigma}^2_{cb} = 33.40$, $\hat{\sigma}^2_{cb} = 24.04$, respectively). The largest component for Factor 2 was coder ($\sigma_{c} = 15.57$).

Generalizability Coefficients

With the large estimates of variance components for the coder x block interaction and the residual at 25 and 43 weeks, one would not expect the generalizability coefficients which estimated the intercoder agreement when generalization was intended to all subjects and coders and both levels of Parent Sex to be very high at those occasions, and such was the case. These coefficients are shown in Table 23 and were derived by substituting the estimates of variance components from Tables 18-22 into the equations shown in Table 6. Sixty different coefficients were generated, four for each factor at each occasion.

The first coefficient at each occasion was $\hat{\rho}^2(s,c,P)$ and was an estimate of the intercoder agreement on a factor score of parent-infant interaction when generalization was intended over all subjects and coders and to both levels of Parent Sex. The only coefficients that are above .70 are for Factors 2, 3, 4 and 5 at 37 weeks. It was therefore necessary to look to the first-order interaction of coder or Parent Sex x block in an attempt to increase the estimate of intercoder agreement for each factor.

At 25 weeks the coefficient $\hat{\rho}^2(s,c,P^*)$ for Factor 5 was .78; at 37 weeks it was .75 for Factor 1. This was an estimate of generalizability when the universe of generalization was to all coders and subjects, but to only one level of Parent Sex. Notice that in order for these two coefficients to become larger than .70, the universe of generalization has become smaller. One could also obtain larger coefficients for Factors 2, 3, 4 and 5 at 37 weeks if one accepted a smaller universe of generalization.

At 43 weeks the coefficient $\hat{\rho}^2(s,c^*,P)$ becomes larger than .70 for all five factors (.84, .83, .77, .91, .78, respectively). This was an estimate of generalizability when the universe of generalization was to all subjects and both levels of Parent Sex but to the specific coders used in this study.

The coefficients for Factors 1 and 3 at 25 weeks did not become larger than .70 until the universe of generalization was restricted to all subjects, to the specific coders used in this study and one level of Parent Sex (.82 and .93, respectively); the coefficients for Factor 2 and 4 never obtained that value (.68 for both).

Reliability Analysis

Subjects Crossed With Coders

Estimates of Variance Components. Each of the factors was again analyzed separately using a randomzied block factorial design. The infants used in this analysis had a mean MDQ of 114.9 and a standard deviation of 11.8. The components of variance for each source of variance for each factor are shown in Tables 24-28. Again, notice that there are negative estimates for each factor, which are especially large for Factors 2, 4, and 5. Again, a majority of these negative estimates involve the facet of Parent Sex. In addition, Factor 4 has a negative estimate for subjects and Factors 4 and 5 have negative estimates for coders.

The large estimated variance component for Factor 1 (parent-infant interaction with baby amplifying) was for Occasion ($\hat{\sigma}_O^2 = 5.10$), followed by the Occasion by subject interaction ($\hat{\sigma}_{Os}^2 = 4.31$, see Table 24). Notice that the estimated variance components for coders and subjects x

Table 24

Components of Variance for Factor 1 (Parent-Infant Interaction; Baby Amplifies) for Reliability Analysis When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	17.90	2	1	.31
Parent Sex	.09	0	1	-.72
Occasion	295.47	26	2	5.10
block	194.02	17	5	2.66
cP	.31	0	1	-.11
cO	24.42	2	2	.42
PO	29.08	3	2	.62
cb	34.50	3	5	1.15
Pb	140.50	12	5	4.31
Ob	202.47	18	10	3.27
cPO	.33	0	2	-.42
cPb	11.27	1	5	.75
cOb	71.84	6	10	3.59
POb	96.72	8	10	3.49
residual	26.95	2	10	2.69
Total	1145.87	100	71	

Table 25

Components of Variance for Factor 2 (Parent-Infant Interaction; Baby Responds) for Reliability Analysis When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	12.58	1	1	.29
Parent Sex	160.51	7	1	3.91
Occasion	339.94	14	2	3.06
block	252.81	11	5	4.04
cP	24.38	1	1	.02
cO	32.58	1	2	.90
PO	34.47	1	2	1.99
cb	10.48	0	5	.35
Pb	96.97	4	5	-.77
Ob	856.73	36	10	20.06
cPO	9.60	0	2	-3.51
cPb	120.08	5	5	8.01
cOb	54.46	2	10	2.72
POb	144.38	6	10	-5.70
residual	258.47	11	10	25.85
Total	2408.44	100	71	

Table 26

Components of Variance for Factor 3 (General Baby Behavior)
for Reliability Analysis When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	89.56	6	1	2.31
Parent Sex	21.23	1	1	-.69
Occasion	435.03	31	2	7.86
block	54.63	4	5	.37
cP	2.92	0	1	-.15
cO	3.20	0	2	-.12
PO	27.42	2	2	.36
cb	32.05	2	5	1.07
Pb	243.43	18	5	7.17
Ob	303.35	21	10	6.80
cPO	3.81	0	2	-.38
cPb	28.48	2	5	1.90
cOb	31.50	2	10	1.57
POb	115.75	8	10	3.71
residual	41.58	3	10	4.16
Total	1433.94	100	71	

Table 27

Components of Variance for Factor 4 (Parent-Infant Interaction; Baby Warms, Accepts) for Reliability Analysis When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	10.97	0	1	-.72
Parent Sex	22.56	1	1	-.58
Occasion	1601.13	49	2	29.31
block	100.90	3	5	-1.40
cP	4.35	0	1	-.38
cO	130.19	4	2	4.43
PO	3.49	0	2	-.88
cb	185.06	6	5	6.17
Pb	251.07	8	5	6.49
Ob	440.64	13	10	8.02
cPO	1.35	0	2	-1.71
cPb	56.22	2	5	3.75
cOb	119.66	4	10	5.98
POb	225.29	7	10	5.79
residual	109.53	3	10	10.95
Total	3262.41	100	71	

Table 28

Components of Variance for Factor 5 (Parent-Infant Interaction; Baby Initiates) for Reliability Analysis When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	3.04	0	1	-1.15
Parent Sex	43.55	1	1	.81
Occasion	1107.28	27	2	18.19
block	436.64	11	5	3.58
cP	18.20	0	1	.77
cO	10.44	0	2	-1.19
PO	83.76	2	2	.39
cb	221.86	5	5	7.40
Pb	295.00	7	5	9.10
Ob	1313.65	31	10	27.95
cPO	12.55	0	2	.18
cPb	22.06	1	5	1.47
cOb	195.54	5	10	9.78
POb	361.04	9	10	15.45
residual	52.09	1	10	5.21
Total	4176.70	100	71	

coders interaction were relatively small ($\hat{\sigma}_{c}^{2} = .31$, $\hat{\sigma}_{cs}^{2} = .62$). The interpretation would be that the ratings of subjects by the two coders were quite similar for each occasion, but the scores and relative rankings did change from occasion to occasion. While it is not proper to test for significant differences with estimates of variance components (Cronbach et al., 1972), it is obvious that the facet Occasion explains much more variance than does the facet coder. This variance does not affect the generalizability coefficient, however, since all subjects were seen by both coders, with both parents and on all three occasions and, therefore, the variance component for these three main effects was not included as part of the observed variance (see Table 9).

The largest estimated variance component for Factor 2 (parent-infant interaction with infant responding) was for the residual followed by the Occasion x subject interaction ($\hat{\sigma}_{residual}^{2} = 25.25$, $\hat{\sigma}_{Os}^{2} = 20.06$, see Table 25). Notice that there are three negative estimates, two of which are larger than -1.00 ($\hat{\sigma}_{Ps}^{2} = -.77$, $\hat{\sigma}_{cPO}^{2} = -3.51$, $\hat{\sigma}_{POs}^{2} = -5.70$). Cronbach et al. (1972) suggest that in cases such as this an incorrect model may be the problem. For example, Parent Sex is represented in each of the six negative estimates for Factors 1 and 2. Parent Sex is also represented in three of the four negative estimates for Factor 3, and four of the six negative estimates for Factor 4. Perhaps a reanalysis of the data using only the facets of coder and Occasion would prove to be a better model (i.e., negative estimates would be a smaller percentage of total number of estimates and/or reduced in value). It is also true that the small number of subjects could be influencing the situation.

The largest estimated variance component for Factor 3 (general baby behavior) was again Occasion ($\hat{\sigma}_O^2$ = 7.86) followed by the Parent Sex x subject interaction ($\hat{\sigma}_{Ps}^2$ = 7.17, see Table 26). Notice that this latter is the case even though the estimated variance component for the main effect of Parent Sex is negative ($\hat{\sigma}_P^2$ = -.69). This means that even though the differences between parents were negligible, the rankings of subjects changed depending upon which parent was interacting with the infant and seems to provide evidence that Parent Sex should not be dropped from the model, at least for this factor. However, if Parent Sex was dropped from the model, the block would become equal to a parent-infant pair so that it would be necessary to be specific about which parent was interacting with the infant.

The largest estimated variance component for Factor 4 (parent-infant interaction with baby warming and/or amplifying) was again for Occasion ($\hat{\sigma}_O^2$ = 29.31, see Table 27). Notice that the main effects for coder, Parent Sex and subjects were each negative. This is especially critical in the case of subjects, since differences between subjects is a necessity for reliable data (Medley and Mitzel, 1963).

The largest estimated variance component for Factor 5 (parent-infant interaction with baby initiating) was for the Occasion x subject interaction followed by Occasion ($\hat{\sigma}_{Os}^2$ = 29.75, $\hat{\sigma}_O^2$ = 18.19, see Table 28). The two negative estimates were for the main effect of coder and the coder x Occasion interaction ($\hat{\sigma}_c^2$ = -1.15, $\hat{\sigma}_{cO}^2$ = -1.19).

Generalizability Coefficients. The generalizability coefficients estimating the reliability of each of the five factors of parent-infant interaction when subjects were crossed with coders are shown in Table 29.

Table 29

Generalizability Coefficients for Reliability Analysis When Subjects Were Crossed With Coders[a]

Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction: Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
$\hat{\rho}^2(b,c,P,O)$	.29	.20	.04	0	.10
$\hat{\rho}^2(b,c^*,P,O)$	.35	.20	.09	.20	.20
$\hat{\rho}^2(b,c,P^*,O)$	.52	.20	.39	.20	.22
$\hat{\rho}^2(b,c,P,O^*)$	.46	.70	.37	.25	.47
$\hat{\rho}^2(b,c^*,P^*,O)$	.60	.30	.49	.46	.33
$\hat{\rho}^2(b,c^*,P,O^*)$	.63	.72	.46	.55	.64
$\hat{\rho}^2(b,c,P^*,O^*)$	.79	.70	.81	.56	.81
$\hat{\rho}^2(b,c^*,P^*,O^*)$	.97	.82	.98	.91	.98

a Substituting 0 for negative estimates of variance components

The coefficients were derived by substituting the estimates of variance components from Tables 24-28 into the equations shown in Table 9. Forty different coefficients were estimated, eight for each factor.

The first coefficient was $\hat{\rho}^2(s,c,P,O)$ and was an estimate of the reliability of a factor score when generalization was intended over all subjects and all coders, both levels of Parent Sex and three levels of Occasion. All of these coefficients were relatively low, with the coefficients for Factors 3, 4 and 5 being especially so. It was therefore necessary to look to a first-order interaction in an attempt to increase the estimate of reliability for each of the factors.

The next three coefficients were $\hat{\rho}^2(s,c^*,P,O)$, $\hat{\rho}^2(s,c,P^*,O)$ and $\hat{\rho}^2(s,c,P,O^*)$ which estimated the reliability of ratings when generalization was intended over all subjects but to only the specific coders in this study, a specific level of Parent Sex or a specific level of Occasion, respectively. Notice that, with only exception, the coefficients are less than .55. The exception is for Factor 2 where the coefficient $\hat{\rho}^2(s,c,P,O^*)$ is .70. For the other factors, then, it was necessary to look to a second-order interaction in order to increase the reliability estimate.

There were again three coefficients which represented second-order interactions. For Factors 1, 3 and 5 the coefficient for $\hat{\rho}^2(s,c,P^*,O^*)$ was larger than .70 (.79, .81, .81, respectively). However, for Factor 4 none of the coefficients exceed .70 and, therefore, it was necessary to look to the third-order coefficient.

This last coefficient was $\hat{\rho}^2(s,c^*,P^*,O^*)$ which was an estimate of reliability of ratings when generalization was intended for all subjects, but to the specific coders used in this study, and to a specific level

of both Parent Sex and Occasion. Factor 4 had a value of .91 for this coefficient.

Subjects Nested Within Coders

Estimates of Variance Components. Each of the factors was again analyzed separately using a five-facet split-plot design. The mean MDQ for Group 1 was 122.6 with a standard deviation of 10.7; for Group 2 the mean was 115.1 with a standard deviation of 13.5. While this difference was not statistically significant ($t = 1.88$, $t_{.025,\ 26} = 2.05$), it was approxiamtely 1/2 standard deviation. The components of variance for each source of variance for each factor are shown in Tables 30-34. Again, notice the large number of negative estimates for each factor. It is interesting that the majority of these negative estimates are interactions involving either coder or infant sex.

The largest estimated variance component for Factor 1 (Parent-infant interaction with baby amplifying) was for the residual ($\hat{\sigma}^2_{residual}$ = 13.95). Since there was only one occasion per cell, the experimental error was confounded with this term and could either magnify or conceal actual differences between coders since coders were nested within groups. For example, the mean for Group 2 was lower than Group 1. If coder 2 were actually assigning lower scores on a particular factor than coder 1, then these differences would be magnified by the differences between groups. However, if coder 2 were assigning higher scores, then the actual differences would be masked.

All of the sources of variance for Factor 1 involving subjects were substantial. This means that there were initial differences between ratings of subjects, but these ratings changed with respect to ranking

Table 30

Components of Variance for Factor 1 (Parent-Infant Interaction; Baby Amplifies) for Reliability Analysis When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	321.59	6	1	2.47
Infant Sex	.22	0	1	-.37
cI	41.50	1	1	-.07
sub/cI	1091.42	19	24	5.68
Parent Sex	.48	0	1	-.42
cP	47.34	1	1	.66
PI	7.00	0	1	.07
cPI	2.96	0	1	-.26
Ps/cI	245.04	4	24	2.55
Age	1032.79	18	1	7.93
cA	144.08	3	1	2.10
IA	109.54	2	1	1.94
cIA	.70	0	1	-.92
As/cI	634.48	11	24	6.61
Task	.01	0	1	-.18
cT	19.81	0	1	-.09
IT	.26	0	1	-.08
cIT	4.50	0	1	-.72
Ts/cI	592.82	10	24	6.18
PA	1.26	0	1	-.18
cPA	11.17	0	1	-.03
PIA	2.02	0	1	-.63
cPIA	19.64	0	1	.54
PAs/cI	288.94	5	24	6.02
PT	20.77	0	1	.24
cPT	7.28	0	1	-.09
PIT	21.26	0	1	.13

Table 30 - continued

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
cPIT	17.62	0	1	.57
PTs/cI	232.60	4	24	4.85
AT	.44	0	1	.01
cAT	.02	0	1	-.56
IAT	38.38	1	1	.60
cIAT	21.58	0	1	.41
ATs/cI	379.44	7	24	7.90
PAT	1.90	0	1	.01
cPAT	2.08	0	1	-.85
PIAT	8.47	0	1	-1.93
cPIAT	35.44	1	1	3.07
residual	334.72	6	24	13.95
Total	5741.60	100	223	

Table 31

Components of Variance for Factor 2 (Parent-Infant Interaction; Baby Responds) for Reliability Analysis When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	77.18	1	1	.38
Infant Sex	17.21	0	1	-.86
cI	113.02	1	1	1.39
sub/cI	842.12	10	24	4.39
Parent Sex	27.93	0	1	.25
cP	.25	0	1	-.41
PI	14.36	0	1	-4.10
cPI	244.21	3	1	7.90
Ps/cI	550.74	6	24	5.74
Age	129.77	2	1	.55
cA	67.67	1	1	.41
IA	135.63	2	1	1.73
cIA	38.84	0	1	-.20
As/cI	1069.70	13	24	11.14
Task	832.66	10	1	5.22
cT	248.53	3	1	3.69
IT	72.66	1	1	-.33
cIT	91.31	1	1	1.76
Ts/cI	1010.28	12	24	10.52
PA	34.66	0	1	.46
cPA	8.66	0	1	-.31
PIA	15.58	0	1	-1.81
cPIA	66.36	1	1	3.51
PAs/cI	414.89	5	24	8.64

Table 31 - continued

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
PT	64.49	1	1	.60
cPT	30.72	0	1	.40
PIT	55.60	1	1	1.69
cPIT	8.40	0	1	-.80
PTs/cI	471.67	6	24	9.83
AT	100.94	1	1	.53
cAT	71.11	1	1	1.19
IAT	18.36	0	1	-.75
cIAT	39.30	0	1	.12
ATs/cI	903.71	11	24	18.83
PAT	1.12	0	1	.02
cPAT	.69	0	1	-1.73
PIAT	7.28	0	1	.53
cPIAT	0	0	1	-3.57
residual	598.20	7	24	24.93
Total	8495.81	100	223	

Table 32

Components of Variance for Factor 3 (General Baby Behavior) for Reliability Analysis When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	47.19	1	1	-.04
Infant Sex	11.34	0	1	-.12
cI	24.84	0	1	-.15
sub/cI	1248.76	23	24	6.50
Parent Sex	31.35	0	1	.18
cP	10.80	0	1	-.18
PI	.18	0	1	-.10
cPI	5.60	0	1	-.54
Ps/cI	497.20	10	24	5.18
Age	1206.20	22	1	9.18
cA	177.87	3	1	2.98
IA	16.95	0	1	.29
cIP	.79	0	1	-.37
As/cI	268.41	5	24	2.80
Task	79.91	1	1	.56
cT	17.39	0	1	-.03
IT	7.88	0	1	-.05
cIT	10.54	0	1	-.30
Ts/cI	451.33	8	24	4.70
PA	6.12	0	1	-.14
cPA	13.79	0	1	.29
PIA	17.37	0	1	.53
cPIA	2.62	0	1	-.21
PAs/cI	133.08	2	24	2.77

Table 32 - continued

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
PT	13.32	0	1	.20
cPT	2.08	0	1	-.23
PIT	.37	0	1	-.01
cPIT	.02	0	1	-.61
PTs/cI	204.73	4	24	4.27
AT	178.95	3	1	2.47
cAT	40.78	1	1	.96
IAT	124.78	2	1	4.46
cIAT	.03	0	1	-1.00
ATs/cI	336.54	6	24	7.01
PAT	8.69	0	1	.13
cPAT	5.18	0	1	-.24
PIAT	6.07	0	1	.10
cPIAT	4.61	0	1	-.56
residual	204.14	4	24	8.51
Total	5417.95	100	223	

Table 33

Components of Variance for Factor 4 (Parent-Infant Interaction; Baby Warms, Accepts) for Reliability Analysis When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	2214.93	16	1	19.18
Infant Sex	50.16	0	1	.43
cI	2.38	0	1	-1.15
sub/cI	1597.76	11	24	8.32
Parent Sex	149.49	1	1	1.10
cP	26.63	0	1	-.14
PI	2.45	0	1	-.14
cPI	10.56	0	1	-.85
Ps/cI	825.81	6	24	8.60
Age	820.04	6	1	6.63
cA	77.74	0	1	.36
IA	57.05	0	1	1.02
cIA	0	0	1	-2.07
As/cI	1386.33	10	24	14.44
Task	734.35	5	1	2.79
cT	421.52	3	1	6.51
IT	9.68	0	1	.15
cIT	1.17	0	1	-1.99
Ts/cI	1362.40	10	24	14.19
PA	29.49	0	1	.52
cPA	.35	0	1	-1.06
PIA	23.86	0	1	-2.72
cPIA	100.11	1	1	5.00
PTs/cI	722.04	5	24	15.04

Table 33 - continued

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
PT	24.66	0	1	.16
cPT	15.52	0	1	-.34
PIT	59.37	0	1	2.03
cPIT	2.60	0	1	-1.60
PTs/cI	599.95	4	24	12.50
AT	551.38	4	1	8.99
cAT	48.21	0	1	.10
IAT	66.21	0	1	1.61
cIAT	21.02	0	1	-1.74
ATs/cI	1087.62	8	24	22.66
PAT	67.19	0	1	2.37
cPAT	.80	0	1	-2.26
PIAT	.27	0	1	-.07
cPIAT	1.24	0	1	-4.47
residual	780.07	6	24	32.50
Total	13952.32	100	223	

Table 34

Components of Variance for Factor 5 (Parent-Infant Interaction; Baby Initiates) for Reliability Analysis When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	43.31	0	1	-.50
Infant Sex	0	0	1	-.01
cI	1.67	0	1	-1.74
sub/cI	2382.67	15	24	12.41
Parent Sex	222.18	1	1	1.95
cP	3.60	0	1	-.62
IP	34.04	0	1	-.05
cIP	31.25	0	1	-.25
Ps/cI	916.53	6	24	9.55
Age	312.76	2	1	.79
cA	224.64	1	1	3.37
IA	77.23	0	1	.84
cIA	30.35	0	1	-.19
As/cI	856.36	5	24	8.92
Task	5155.74	32	1	45.96
cT	8.22	0	1	-.95
IT	23.97	0	1	.31
cIT	6.44	0	1	-1.97
Ts/cI	1476.03	9	24	15.38
PA	7.52	0	1	-.18
cPA	17.49	0	1	-.90
PIA	2.80	0	1	-4.62
cPIA	132.28	1	1	6.39
PAs/cI	1027.70	6	24	21.41

Table 34 - continued

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
PT	58.93	0	1	.96
cPT	5.16	0	1	-.56
PIT	0	0	1	-1.34
cPIT	37.13	0	1	1.16
PTs/cI	500.16	3	24	10.42
AT	711.50	4	1	10.27
cAT	136.45	1	1	3.36
IAT	30.40	0	1	-6.82
cIAT	221.43	1	1	12.80
ATs/cI	1014.16	6	24	21.13
PAT	0	0	1	-.13
cPAT	3.17	0	1	-1.50
PIAT	3.20	0	1	-.92
cPIAT	16.04	0	1	-1.16
residual	580.44	4	24	24.19
Total	16312.95	100	223	

of subjects dependent upon the level of facet or facets which were included in the study. This does not necessarily mean there were differences between levels of the facet, but it does not preclude differences either. For example, the estimate of the variance component for Type of Task x subject(coder x Infant Sex) was almost equal to that for Age of Infant x subject(coder x Infant Sex) ($\hat{\sigma}^2_{Ts(cI)} = 6.18$, $\hat{\sigma}^2_{As(cI)} = 6.61$). Yet the estimate for Type of Task ($\hat{\sigma}^2_T = -.18$) indicated no differences between task levels, while the estimate for Age of Infant ($\hat{\sigma}^2_A = 7.93$) indicated substantial differences between infant age levels. This latter variance does not affect the generalizability coefficient, however, since all subjects were seen at both ages and, therefore, the variance component for Infant Age Grouping was not included as part of the observed score variance (see Table 12).

The estimates of variance components for Factor 2 (Parent-infant interaction with baby responding) showed a very similar pattern. Again, the largest estimate is for the residual (residual = 24.93) followed by the Infant Age Grouping x Type of Task x subjects(coder x Infant Sex) interaction ($\hat{\sigma}_{ATs(cI)} = 18.83$). Also, all of the estimates involving subjects each accounted for at least 5% of the total sums of squares.

For Factor 3 (General baby behavior) the largest estimate of variance component was for Infant Age Grouping ($\hat{\sigma}^2_A = 9.18$). The second and third largest were for the residual ($\hat{\sigma}_{residual}^2 = 8.51$), and the Infant Age Grouping x Type of Task x subjects(coder x Infant Sex) interaction ($\hat{\sigma}^2_{ATs(cI)} = 7.01$). Again, with only two exceptions, each of the sources of variance involving subjects accounted for at least 5% of the total sums of squares.

The estimates of variance components for Factor 4 (Parent-infant interaction with baby warming or accepting) repeated the, by now familiar, pattern ($\hat{\sigma}^2_{residual} = 32.50$, $\hat{\sigma}^2_{ATs(cI)} = 22.66$). However, for Factor 4 the estimate for the main effect of coder accounted for 16% of the total sums of squares ($\hat{\sigma}^2_c = 19.18$). It should be emphasized that in the split-plot design variance due to differences between coders and variance due to differences between groups to which coders were assigned was completely confounded, so that it is possible that this estimate was overestimating differences between coders.

For Factor 5 (Parent-infant interaction with baby initiating) the largest estimate of variance component was for the main effect of Type of Task ($\hat{\sigma}^2_T = 45.96$). The next three largest were for the residual, the Infant Age Grouping x Type of Task x subject(coder x Infant Sex) interaction and the Parent Sex x Infant Grouping x subjects(coders x Infant Sex) interaction ($\hat{\sigma}^2_{residual}$ 24.19, $\hat{\sigma}^2_{ATs} = 21.13$, $\sigma^2_{PAs(cI)} =$ 21.44).

To summarize the information about the estimates of variance components, it was noted that there w as a large number of negative estimates which were obviously poor estimates since a variance is, by definition, non-negative. The largest estimates were consistently those associated with the residual and the Infant Age Grouping x Type of Task x subjects(coders x Infant Sex) interaction.

Generalizability Coefficients. The generalizability coefficients estimating the reliability of each of the five factors of parent-infant interaction are shown in Table 35. The coefficients were derived by substituting the estimates of variance components from Tables 30-34 into

Table 35

Generalizability Coefficients for Reliability Analysis When Subjects Were Nested Within Coders[a]

Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
$\hat{\rho}^2(s,c,I,P,A,T)$	.25	.13	.34	.13	.26
$\hat{\rho}^2(s,c,I,P^*,A,T)$	.25	.21	.48	.19	.36
$\hat{\rho}^2(s,c,I,P,A^*,T)$	.40	.29	.41	.24	.35
$\hat{\rho}^2(s,c,I,P,A,T^*)$	.39	.28	.46	.24	.42
$\hat{\rho}^2(s,c,I,P^*,A^*,T)$	.47	.44	.58	.36	.56
$\hat{\rho}^2(s,c,I,P^*,A,T^*)$	.44	.43	.65	.49	.57
$\hat{\rho}^2(s,c,I,P,A^*,T^*)$	.62	.58	.63	.43	.62
$\hat{\rho}^2(s,c,I,P^*,A^*,T^*)$	.75	.80	.86	.74	.88

a Substituting 0 for negative estimates of variance components

the equations shown in Table 12. Forty different coefficients were estimated, eight for each factor.

The first coefficient was $\hat{\rho}^2(s,c,I,P,A,T)$ and was an estimate of the reliability of a factor when generalization was intended over all of the facets used in this study. All of these were relatively low, the largest being .34 for Factor 3. It was therefore necessary to look to the first-order interactions of facets by subjects in an attempt to increase the reliability estimates for each of the factors. This was done at the expense of generalizing to a smaller universe.

The next three generalizability coefficients were $\hat{\rho}^2(s,c,I,P^*,A,T)$, $\hat{\rho}^2(s,c,I,P,A^*,T)$ and $\hat{\rho}^2(s,c,I,P,A,T^*)$, which estimated the reliability of ratings when generalization was not intended to levels of Parent Sex, Infant Age Grouping or Type of Task, respectively. None was higher than .50, the largest being .48 for Factor 3. If increased reliability were desired one could further reduce the universe to which generalization would be made by looking at the second-order interactions of facets by subjects.

Accordingly, the next three generalizability coefficients were $\hat{\rho}^2(s,c,I,P^*,A^*,T)$, $\hat{\rho}^2(s,c,I,P^*,A,T^*)$ and $\hat{\rho}^2(s,c,I,P,A^*,T^*)$ which estimated the reliability of ratings when generalization was not intended to levels of Parent Sex or Infant Age Grouping, Parent Sex or Type of Task, and Infant Age Grouping or Type of Task, respectively. For example, if one were concerned with the reliability of each factor for a given taping session one would look at the generalizability coefficient $\hat{\rho}^2(s,c,I,P,A^*,T^*)$ which was .62 for Factor 1, .58 for Factor 2, etc.

If still increased reliability were desired one could look at the last coefficient generated ($\hat{\rho}^2$(s,c,I,P*,A*,T*)), which estimated the reliability of ratings when generalization was only intended for subjects, coders and Infant Sex. The estimate of this coefficient was conservative since the variance due to PATs(cI) was confounded with experimental error and, therefore, not included as part of the estimate of the universe score. Nevertheless, the coefficients ranged from a low of .75 and .74 (for Factors 1 and 4, respectively) to a high of .88 (for Factor 5), which would seem to be adequate for most decision studies.

Additional Analyses

Intercoder Agreement

Since a large number of the estimates of variance components for Parent Sex in the original analysis of intercoder agreement were negative, a second analysis was done which eliminated that facet. The full model for the randomized block design used in this analysis is the same as that shown in Table 4 with P_j removed. Again, coders and blocks were considered random; a block in this model indicates a parent-infant pair. The algorithms for computing the Expected Mean Squares and the Estimates of Variance Components are the same as shown in Table 8 with all reference to Parent Sex likewise removed; in this case the coder x block interaction (cb) is the residual.

The components of variance for each factor are shown in Tables 36-40. Using this model there were only five negative estimates (four involving the main effect of coder) and only one was above 1.0. For all factors at 37 weeks the largest estimated variance component was for block. However, at 25 and 43 weeks the largest estimated variance component for nine of ten cases was for the coder x block interaction.

Table 36

Components of Variance for Additional Intercoder Agreement Analysis of Factor 1 (Parent-Infant Interaction; Baby Responds)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	126.58	19	1	4.23
	block	306.02	47	27	1.60
	cb	219.71	34	27	8.14
	Total	652.31	100	55	
37 weeks	coder	1.74	0	1	-.17
	block	1118.09	80	35	12.11
	cb	270.41	19	35	7.72
	Total	1390.28	99	71	
43 weeks	coder	23.97	2	1	.86
	block	759.83	56	33	3.00
	cb	560.92	42	33	17.00
	Total	1345.92	100	67	

Table 37

Components of Variance for Additional Intercoder Agreement Analysis of Factor 2 (Parent-Infant Interaction; Baby Responds)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	51.04	2	1	.92
	block	1541.09	68	27	15.87
	cb	683.93	30	27	25.33
	Total	2276.06	100	55	
37 weeks	coder	22.76	1	1	.37
	block	1575.11	82	35	17.83
	cb	326.73	17	35	9.34
	Total	1924.60	100	71	
43 weeks	coder	549.49	29	1	1.80
	block	879.05	46	33	5.91
	cb	488.34	25	33	14.80
	Total	1916.88	100	67	

Table 38

Components of Variance for Additional Intercoder Agreement Analysis of Factor 3 (General Baby Behavior)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	6.71	1	1	.04
	block	441.68	73	27	5.35
	cb	153.02	25	27	5.67
	Total	601.41	99	55	
37 weeks	coder	119.08	8	1	3.20
	block	1224.43	83	35	15.58
	cb	133.70	9	35	3.82
	Total	1477.21	100	71	
43 weeks	coder	49.13	3	1	.97
	block	945.84	62	33	6.29
	cb	530.85	35	33	16.09
	Total	1525.82	99	67	

Table 39

Components of Variance for Additional Intercoder Agreement Analysis of Factor 4 (Parent-Infant Interaction; Baby Warms, Accepts)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	13.84	1	1	-.60
	block	1050.37	55	27	4.11
	cb	828.50	44	27	30.69
	Total	1892.71	100	55	
37 weeks	coder	199.00	6	1	5.22
	block	2584.55	82	35	31.45
	cb	383.14	12	35	10.96
	Total	3166.69	100	71	
43 weeks	coder	151.82	7	1	3.43
	block	975.74	42	33	-2.85
	cb	1163.46	51	33	35.26
	Total	2291.02	100	67	

Table 40

Components of Variance for Additional Intercoder Agreement Analysis of Factor 5 (Parent-Infant Interaction; Baby Initiates)

Occasion	Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
25 weeks	coder	13.46	0	1	-.26
	block	2608.93	82	27	37.98
	cb	558.11	18	27	20.67
	Total	3180.50	100	55	
37 weeks	coder	.09	0	1	-.12
	block	2262.93	94	35	30.17
	cb	151.28	6	35	4.32
	Total	2414.30	100	71	
43 weeks	coder	192.48	9	1	4.83
	block	941.85	46	33	.07
	cb	937.25	45	33	28.40
	Total	2071.59	100	67	

The intercoder agreement for ratings of factor scores of parent-infant interaction for each session was evaluated through the calculation of a generalizability coefficient which is shown in Table 41. Again, this coefficient was not an estimate of reliability because variance due to occasion was not considered. Notice that by eliminating the facet of Parent Sex from the analysis, we are now only able to generate one generalizability coefficient. This coefficient is approximately equal to the coefficient $\hat{\rho}^2(b,c,P^*)$ shown in Table 23. The reason that only one coefficient can be generated is that the variance due to the coder x block interaction is now confounded with that due to experimental error; also, the variance due to the Parent Sex x block interaction is now a part of the variance due to block.

The generalizability coefficient for Factor 1 (parent-infant interaction with baby amplifying) at 25 weeks was .28. This increased to .76 at 37 weeks and then declined to .26 at 43 weeks. An examination of the components of variance indicates that the problem at 25 weeks was largely a result of little variance between blocks, while at 43 weeks the problem was more a function of differences between coders.

The intercoder agreement for Factor 2 (Parent-infant interaction with baby responding) showed a slightly different situation. Again, the generalizability coefficient was best at 37 weeks (.79), while considerably less at 25 and 43 weeks (.56 and .44, respectively). However, in this case the variance due to blocks remained steady from 25 weeks to 37 weeks; it then dropped dramatically at 43 weeks. The fact that the variance due to the coder x block interaction fluctuated widely over the three occasions indicates some problem in intracoder stability.

Table 41

Generalizability Coefficients for Additional Intercoder Agreement Analysis[a]

Occasion	Coefficient Name	Description	Algorithm[b]	Generalizability Coefficient for Factors 1	2	3	4	5
25 weeks	$\hat{\rho}^2(b,c)$	Generalization intended to all parent-infant pairs and all coders	$\frac{\hat{\sigma}_b^{\,2}}{\hat{\sigma}_x^{\,2}}$	.28	.56	.65	.21	.79
37 weeks	$\hat{\rho}^2(b,c)$	Generalization intended to all parent-infant pairs and all coders	$\frac{\hat{\sigma}_b^{\,2}}{\hat{\sigma}_x^{\,2}}$	.76	.79	.89	.85	.93
43 weeks	$\hat{\rho}^2(b,c)$	Generalization intended to all parent-infant pairs and all coders	$\frac{\hat{\sigma}_b^{\,2}}{\hat{\sigma}_x^{\,2}}$	.26	.44	.44	.0	.0

a See Table 7 for explanation of notation

b $\hat{\sigma}_x^{\,2} = \hat{\sigma}_b^{\,2} + \frac{1}{Q}\hat{\sigma}_{cb}^{\,2}$

The analysis of intercoder agreement for Factor 3 (General baby behavior) shows still a different situation. In this case the generalizability coefficients for 25 weeks and 43 weeks were both moderate (.65 and .44, respectively), while the coefficient at 37 weeks was much higher (.89). Notice that the problem with differences between subjects was similar to that for Factor 1 but the variance due to different ratings for subjects increased dramatically at 43 weeks. This factor was composed only of measures of baby behavior and is, therefore, some indication of the lack of parent verbal behavior. It is possible, therefore, that one coder was not coding as much parent verbal behavior as the other coder.

The analysis of intercoder agreement for Factor 4 (Parent-infant interaction with baby warming, accepting) showed a pattern similar to Factor 1. The generalizability coefficient for 25 weeks was low (.21), it increased at 37 weeks (.85) and then decreased at 43 weeks (0). The lack of variance between subjects and the wide variation in the coders' ratings of subjects again both seem to be problems. Notice that at 43 weeks the variance due to differences between subjects decreased by 62% while the variance due to ratings by coders for subjects increased over 200% resulting in a generalizability coefficient of 0.

The analysis of intercoder agreement for Factor 5 (Parent-infant interaction with baby initiating) showed a pattern similar to that of Factor 2 in that the variance due to blocks remained stable from 25 to 37 weeks while the variance due to different ratings by coders decreased, resulting in a larger generalizability coefficient (.79 and .93, respectively). However, in this case the variance due to differences between

subjects decreased dramatically at 43 weeks, while the variance due to ratings by coders shifted dramatically from occasion to occasion, resulting again in a generalizability coefficient of 0.

In summary, there were apparently some severe problems with respect to intercoder agreement at both 25 and 43 weeks. As stated previously, the lack of variance between subjects and the lack of intracoder stability seem to be the major sources of the problems. Only Factor 5 showed an acceptable level at either of these two occasions.

All of the generalizability coefficients show improvement at 37 weeks. From an analysis of the means for the coder facet, it would appear that there was a systematic bias on the part of coder 1 to assign higher scores for Factor 3 and lower scores for Factor 4, yet both generalizability coefficients were larger than .80. This could happen if both coders were ranking subjects in a similar manner even though one was consistently assigning higher scores.

Reliability

Subjects crossed with coders. The original reliability analysis for subjects crossed with coders also had a large number of negative estimates of variance components for the facet of Parent Sex, and, therefore, a second analysis was done. The full model for the randomized block factorial design used in this analysis is the same as that shown in Table 7 with P_j removed; a block in this model indicates a parent-infant pair. The algorithms for computing the Expected Mean Squares and the Estimates of the Variance Components are the same as shown in Table 8 with all references to Parent Sex removed; in this case the coder x Occasion x block interaction (cOb) is the residual.

The components of variance for each factor are shown in Tables 42-46. Using this model there were only five negative estimates (three involving the main effect of coder and two involving the coder x Occasion interaction), with the largest being -.58. The largest estimates for Factors 1, 3 and 5 were for either the main effect of Occasion or the Occasion x block interaction; for Factor 2 they were the Occasion x block interaction and the residual; for Factor 4 they were the main effect of Occasion and the residual. The algorithms for computing the generalizability coefficients for this design are shown in Table 47.

None of the factors had a value above .70 for the coefficient $\hat{\rho}^2$ (b,c,O) so it was necessary to look to a first-order interaction (see Table 48). Factors 1, 3 and 5 had values above .70 for the coefficient $\hat{\rho}^2$ (b,c,O*) which was an estimate of reliability when generalization was intended over all parent-infant pairs and coders, but to only one level of Occasion (.79, .75, .78, respectively). It was not until generalization was restricted to all parent-infant pairs, this specific set of coders and one level of Occasion that the generalizability coefficient was above .70 for Factors 2 and 4 (.81, and .80, respectively).

Subjects nested within coders. The original reliability analysis for subjects crossed with coders had a large number of negative estimates of variance components for the facets of Infant Sex and Parent Sex, and, therefore, these were eliminated from the model and a second analysis was done. The full model for the three-facet split-plot design used in this analysis is the same as that shown in Table 10 with I_k and P_j removed; subjects in this model refers to a parent-infant pair. The algorithms for computing the Expected Mean Squares and the Estimates of

Table 42

Components of Variance for Additional Reliability Analysis of Factor 1 (Parent-Infant Interaction; Baby Amplifying) When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coder	17.90	1	1	.38
Occasion	295.47	26	2	5.21
blocks	334.60	29	11	4.37
cO	24.42	2	2	.64
cb	46.08	4	11	1.40
Ob	328.28	29	22	5.21
residual	99.12	9	22	4.51
Total	1145.87	100	71	

Table 43

Components of Variance for Additional Reliability Analysis of Factor 2 (Parent-Infant Interaction; Baby Responds) When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	12.58	1	1	-.04
Occasion	339.94	14	2	5.05
blocks	510.29	21	11	5.38
cO	32.58	2	2	.14
cb	154.96	6	11	4.69
Ob	1035.59	43	22	16.21
residual	322.52	13	22	14.66
Total	2408.46	100	71	

Table 44

Components of Variance for Additional Reliability Analysis of Factor 3 (General Baby Behavior) When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	89.56	6	1	2.33
Occasion	435.03	31	2	8.30
blocks	319.30	22	11	3.88
cO	3.22	0	2	-.16
cb	63.44	5	11	1.92
Ob	446.52	31	22	8.40
residual	76.88	5	22	3.49
Total	1433.95	100	71	

Table 45

Components of Variance for Additional Reliability Analysis of Factor 4 (Parent-Infant Interaction; Baby Warms, Accepts) When Subjects Were Crossed With Coder

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	10.97	0	1	-.32
Occasion	1601.13	49	2	29.81
blocks	374.53	11	11	1.95
cO	130.18	4	2	4.55
cb	245.63	8	11	7.44
Ob	669.41	21	22	9.97
residual	230.53	7	22	10.48
Total	3262.38	100	71	

Table 46

Components of Variance for Additional Reliability Analysis of Factor 5 (Parent-Infant Interaction; Baby Initiates) When Subjects Were Crossed With Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	3.04	0	1	-.58
Occasion	1107.28	27	2	20.01
blocks	775.20	19	11	7.77
cO	10.44	0	2	-.55
cb	262.12	6	11	7.94
Ob	1758.45	42	22	34.05
residual	260.18	6	22	11.83
Total	4176.71	100	71	

Table 47

Description and Algorithms for Additional Analysis of Generalizability When Subjects Were Crossed With Coders[a]

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficient[b]
$\hat{\rho}^2(b,c,O)$	Generalization intended over all parent-infant pairs and coders and to all levels of Occasion included in this analysis	$\dfrac{\hat{\sigma}_b^{\,2}}{\hat{\sigma}_x^{\,2}}$
$\hat{\rho}^2(b,c^*,O)$	Generalization intended over all parent-infant pairs and to all levels of Occasion used in this analysis, but to the specific coders used in this study	$\dfrac{\hat{\sigma}_b^{\,2} + \frac{1}{Q}\hat{\sigma}_{cb}^{\,2}}{\hat{\sigma}_x^{\,2}}$
$\hat{\rho}^2(b,c,O^*)$	Generalization intended over all parent-infant pairs and coders, but to only one level of Occasion	$\dfrac{\hat{\sigma}_b^{\,2} + \frac{1}{T}\hat{\sigma}_{Ob}^{\,2}}{\hat{\sigma}_x^{\,2}}$
$\hat{\rho}^2(b,c^*,O^*)$	Generalization intended over all parent-infant pairs, but to only one level of Occasion and to the specific coders used in this study	$\dfrac{\hat{\sigma}_b^{\,2} + \frac{1}{Q}\hat{\sigma}_{cb}^{\,2} + \frac{1}{T}\hat{\sigma}_{Ob}^{\,2}}{\hat{\sigma}_x^{\,2}}$

a See explanation of notation in Table 6

b $\hat{\sigma}_b^{\,2} = \hat{\sigma}_b^{\,2} + \frac{1}{Q}\hat{\sigma}_{cb}^{\,2} + \frac{1}{T}\hat{\sigma}_{Ob}^{\,2} + \frac{1}{QT}\hat{\sigma}_{residual}^{\,2}$

Table 48

Generalizability Coefficients for Additional Reliability Analysis When Subjects Were Crossed With Coders[a]

Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
$\hat{\rho}^2(b,c,0)$	.50	.28	.36	.15	.24
$\hat{\rho}^2(b,c^*,0)$	.50	.40	.45	.43	.37
$\hat{\rho}^2(b,c,0^*)$	.79	.69	.75	.52	.78
$\hat{\rho}^2(b,c^*,0^*)$	.80	.81	.84	.80	.91

a Substituting 0 for negative estimates of variance

Variance Components are the same as shown in Table 11 with all reference to Infant Sex and Parent Sex removed; in this case ATs(c) is the residual.

The components of variance for each factor are shown in Tables 49-53. Using this model there were only seven negative estimates with the largest being -.79. The largest estimate for each factor was for the residual with the exception of Factor 5 where the largest estimate was for the main effect of Type of Task.

The algorithm for computing the generalizability coefficients for this design are shown in Table 54. Notice that since all subjects were not observed by all coders the variance due to coders was included as part of the observed score variance. In this case variance due to coders is confounded with variance due to groups and, therefore, the estimates of the generalizability coefficients will be conservative. None of the coefficients had a value above .70 for the coefficient $\hat{\rho}^2(s,c,A,T)$ so it was necessary to look to a first-order interaction (see Table 55). None of the factors had a value above .70 for either of the two first-order interaction coefficients so that it was necessary to look to the second-order interaction. Factors 3 and 5 had values of .78 and .73, respectively, for the coefficient $\hat{\rho}^2(s,c,A^*,T^*)$, but the value of the coefficients for Factors 1, 2, and 4 did not reach .70 (.66, .63 and .59, respectively).

Summary

Thirty-two measures of parent-infant interaction from the Reciprocal Category System were factor analyzed and rotated to the Varimax criterion. The resulting five factors accounted for 75% of the common variance and were basically determined by variance due to the babies' behavior.

Table 49

Components of Variance for Additional Reliability Analysis of Factor 1 (Parent-Infant Interaction; Baby Amplifies) When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	321.88	6	1	2.72
Age	1032.78	18	1	7.93
Task	.01	0	1	-.18
sub/c	919.59	16	54	4.26
cA	144.55	3	1	2.28
cT	20.08	0	1	-.05
AT	.46	0	1	-.02
As/c	909.41	16	54	8.42
Ts/c	1223.44	21	54	11.33
cAT	0	0	1	-.79
residual	1170.08	20	54	21.67
Total	5742.28	100	223	

Table 50

Components of Variance for Additional Reliability Analysis of Factor 2 (Parent-Infant Interaction; Baby Responds) When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	77.31	1	1	.44
Age	129.77	1	1	.55
Task	832.68	10	1	5.21
sub/c	1526.82	18	54	7.07
cA	67.63	1	1	.79
cT	248.62	3	1	3.85
AT	100.93	1	1	.53
As/c	1274.79	15	54	11.80
Ts/c	1779.72	21	54	16.48
cAT	71.08	1	1	.96
residual	2386.50	28	54	44.19
Total	8495.85	100	223	

Table 51

Components of Variance for Additional Reliability Analysis of Factor 3 (General Baby Behavior) When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	47.06	1	1	.22
Age	1206.16	22	1	9.18
Task	79.91	2	1	.56
sub/c	1215.30	22	54	5.63
cA	178.29	3	1	2.92
cT	17.28	0	1	-.03
AT	178.99	3	1	2.47
As/c	801.59	15	54	7.42
Ts/c	1032.33	19	54	9.48
cAT	40.72	1	1	1.04
residual	629.50	12	54	11.66
Total	5418.13	100	223	

Table 52

Components of Variance for Additional Reliability Analysis of Factor 4 (Parent-Infant Interaction; Baby Warms, Accepts) When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	2215.57	16	1	19.40
Age	820.04	6	1	6.63
Task	734.30	5	1	2.79
sub/c	2310.29	17	54	10.70
cA	77.77	1	1	.69
cT	422.17	3	1	6.74
AT	551.41	4	1	9.00
As/c	2121.97	15	54	19.65
Ts/c	2404.57	17	54	22.26
cAT	47.51	0	1	.21
residual	2246.65	16	54	41.60
Total	13952.25	100	223	

Table 53

Components of Variance for Additional Reliability Analysis of Factor 5 (Parent-Infant Interaction; Baby Initiates) When Subjects Were Nested Within Coders

Source	Sums of Squares	% of Total SS	Degrees of Freedom	Estimates of Variance Components
coders	43.33	0	1	-.01
Age	312.76	2	1	.79
Task	5155.81	32	1	45.96
sub/c	2430.37	15	54	11.25
cA	224.46	1	1	3.35
cT	8.21	0	1	-.79
AT	711.42	4	1	10.27
As/c	1982.70	12	54	18.36
Ts/c	2839.06	17	54	26.29
cAT	136.34	1	1	3.24
residual	2467.49	15	54	45.69
Total	16312.06	99	223	

Table 54

Description and Algorithms for Additional Analysis of Generalizability When Subjects Were Nested Within Coders[a]

Generalizability Coefficient	Description	Algorithm for Estimation of Generalizability Coefficient[b]
$\hat{\rho}^2(s,c,A,T)$	Generalization intended over all subjects (parent-infant pairs), all coders and to all levels of Age and Task included in this analysis	$\dfrac{\hat{\sigma}_{s(c)}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(s,c,A^*,T)$	Generalization intended over all subjects and coders, all levels of Task included in the analysis, but to a specific level of Age	$\dfrac{\hat{\sigma}_{s(c)}^2 + \frac{1}{O}\hat{\sigma}_{As(c)}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(s,c,A,T^*)$	Generalization intended over all subjects and coders, and all levels of Age included in the analysis, but to a specific level of Task	$\dfrac{\hat{\sigma}_{s(c)}^2 + \frac{1}{W}\hat{\sigma}_{Ts(c)}^2}{\hat{\sigma}_x^2}$
$\hat{\rho}^2(s,c,A^*,T^*)$	Generalization intended over all subjects and coders, but to a specific level of Age and a specific level of Task	$\dfrac{\hat{\sigma}_{s(c)}^2 + \frac{1}{O}\hat{\sigma}_{As(c)}^2 + \frac{1}{W}\hat{\sigma}_{Ts(c)}^2}{\hat{\sigma}_x^2}$

a See explanation of notation in Table 10

b $\hat{\sigma}_x^2 = \hat{\sigma}_{s(c)}^2 + \frac{1}{O}\hat{\sigma}_{As(c)}^2 + \frac{1}{W}\hat{\sigma}_{Ts(c)}^2 + \frac{1}{OW}\hat{\sigma}_{residual}^2 + \frac{1}{V}\hat{\sigma}_c^2 + \frac{1}{VO}\hat{\sigma}_{cA}^2 + \frac{1}{VW}\hat{\sigma}_{cT}^2 + \frac{1}{VOW}\hat{\sigma}_{cAT}^2$

Table 55

Generalizability Coefficients for Additional Reliability Analysis When Subjects Were Nested Within Coders[a]

Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
$\hat{\rho}^2(s,c,A,T)$	.20	.21	.30	.20	.24
$\hat{\rho}^2(s,c,A^*,T)$	.39	.38	.52	.38	.44
$\hat{\rho}^2(s,c,A,T^*)$	.46	.45	.58	.44	.53
$\hat{\rho}^2(s,c,A^*,T^*)$	.66	.63	.78	.59	.73

a Substituting 0 for negative estimates of variance

Three separate analyses were done: one for intercoder agreement and two for reliability. In each of these analyses there were negative estimates associated with either Parent Sex or Infant Sex or both. Therefore, three additional analyses were done which eliminated these two facets from the models.

In the second analysis of intercoder agreement (where a block equaled a parent-infant pair) it was shown that in nine of ten cases for 25 and 43 weeks the largest estimated variance component was for the coder x block interaction. However, at 37 weeks the largest estimated variance component for all cases was for block. Consequently, the generalizability coefficients estimating intercoder agreement at 37 weeks were all above .70. However, at 25 and 43 weeks only the coefficient for Factor 5 at 25 weeks was above .70.

In the second analysis of reliability when blocks (a block equaled a parent-infant pair) were crossed with coders it was shown that the largest estimates of variance components for Factors 1, 3 and 5 were for either the main effect of Occasion or the Occasion x block interaction; for Factor 2 they were the Occasion x block interaction and the residual; for Factor 4 they were the main effect of Occasion and the residual. For Factors 1, 3 and 5 the values of the generalizability coefficients were above .70 when generalization was intended to all parent-infant pairs and all coders, but to only one level of Occasion (.79, .75 and .78, respectively). For Factors 2 and 4 the values of the generalizability coefficients were larger than .70 only when generalization was intended to all parent-infant pairs, but to the specific coders used in this analysis and one specific level of Occasion (.81 and .80, respectively).

In the second analysis of reliability when subjects (a subject being a parent-infant pair) were nested within coders it was shown that the largest estimate of variance component for each factor was for the residual with the exception of Factor 5 where the largest estimate was for the main effect of Type of Task. It was pointed out that in this analysis the variance due to coders was confounded with the variance due to group so that the estimates of the generalizability coefficients would be conservative. Nevertheless, the coefficients for Factors 3 and 5 were above .70 when generalization was intended to all parent-infant pairs, but to a specific level of Age of Infant and a specific level of Type of Task (.78 and .73, respectively). The coefficients for Factors 1 and 2 were .66 and .63, respectively while the coefficient for Factor 4 was .59.

CHAPTER V

DISCUSSION AND CONCLUSIONS

The purpose of this study was to investigate the generalizability of factors of parent-infant interaction. There were two major aspects: first, the generation of factor scores; second, the analysis of the generalizability of those factor scores.

Factors of Parent-Infant Interaction

Five factors were derived from 32 measures obtained by systematic observation of video-tapes using the Reciprocal Category System (RCS). These measures accounted for 84% of the total tallies recorded by the observers and were an elaboration of two parent-infant patterns reported by Gordon (1975). The first pattern was called Ping-Pong and was described as looking like a Ping-Pong game: Parent does something, baby does something, parent does something, baby does something. Four of the five factors derived in this project fit the Ping-Pong pattern. On Factor 1 the baby is amplifying; on Factor 2 the baby is responding; on Factor 4 the baby is warming or accepting; and on Factor 5 the baby is initiating. On each of these factors all of the parent variables that were used (accepts, amplifies, elicits, initiates and directs) loaded together on the same factor dependent upon the baby behavior. This confirms previous work by Gordon (1974), Barkow et al. (1973) and Clark-Stewart (1973), which indicated that parent behaviors tend to be highly correlated and therefore cluster together.

A second pattern identified by Gordon (1975) was called Persistence and was described as the child being permitted or encouraged to continue an activity on his or her own. Factor 3, which was labeled General Baby Behavior, corresponds to this second pattern.

A third pattern was called Professor and described as parent talking followed by parent talking without paying attention to whether or not anyone is tuned in. However, this pattern was not seen with enough frequency in the present project to be considered. All of the parents in this project seemed to be particularly attentive to their infants. This was probably due to differences in the two types of projects. In the earlier project (Gordon and Jester, 1972) low-income mothers were being taught to play with their infants by parent educators, whereas in the present project (Gordon and Soar, in progress) middle-class mothers and fathers were playing with their infants alone.

Generalizability Analysis

Designs

To date there have been no published studies by researchers using systematic observation to study parent-infant interaction which have reported more than intercoder agreement coefficients as measures of reliability even though that method has been shown to be inappropriate (see Medley and Mitzel, 1963; Cronbach et al., 1963). The purpose of this study was to use three separate designs to investigate the generalizability of factor scores of parent-infant interaction.

The first design was a randomzied block factorial design with the two facets being coder and Parent Sex and a block designated as an infant. The data were analyzed at three separate occasions: 25 weeks,

37 weeks, and 43 weeks. Because of the large number of negative estimates of variance components for Parent Sex, a second analysis was done with the single facet of coder and a block designated as a parent-infant pair. In this second analysis there were only five negative estimates out of the 45 produced, four of which involved the main effect of coder.

The generalizability coefficients were estimates of intercoder agreement and are shown in Table 41. The value of the coefficients was above .70 for all factors at 37 weeks. However, at 25 and 43 weeks only the value for Factor 5 was above .70.

The second design was again a randomized block factorial design which provided an analysis of reliability of factor scores of parent-infant interaction when subjects were crossed with coders. The three facets were coders, Parent Sex and Occasion and a block was designated as an infant. There were again a large number of negative estimates of variance components associated with the facet of Parent Sex, so that facet was eliminated and a second analysis done. Using this second model there were only five negative estimates out of the 35 produced, three involving the main effect of coder and two involving the coder x Occasion interaction. The three Occasions used were again 25 weeks, 37 weeks and 43 weeks.

The generalizability coefficients were estimates of reliability and are shown in Table 48. Factor 1 (Parent-infant interaction with baby amplifying), Factor 3 (General Baby Behavior) and Factor 5 (Parent-infant interaction with baby initiating) all had values above .70 for the coefficient $\hat{\rho}^2$ (b,c,O*), which was an estimate of reliability when generalization was intended over all parent-infant pairs and all coders,

but to a specific level of Occasion. Factor 2 (Parent-infant interaction with baby responding) and Factor 4 (Parent-infant interaction with baby warming, accepting) had values above .70 for the coefficient $\hat{\rho}^2(b,c^*,O^*)$, which was an estimate of reliability when generalization was intended over all parent-infant pairs, but to the specific coders used in this study and to one level of Occasion. However, Factor 2 had a value of .69 for the coefficient $\hat{\rho}^2(b,c,O^*)$, so that it was really only Factor 4 which seemed to have substantial problems with reliability.

The third design was a five-facet split-plot design using the facets of coder, Infant Sex, Parent Sex, Age of Infant and Type of Task. This design provided an analysis of reliability of factor scores when subjects (designated as infants) were nested within coders. Again, there were a large number of negative estimates of variance components and a second analysis was done using only the facets of coder, Age of Infant and Type of Task; subject in this model was designated as a parent-infant pair. In this second analysis there were seven negative estimates out of the 55 produced with only two being larger in absolute value than -.20.

It was anticipated that by partitioning Occasion into two separate facets, one of Age and one of Task difficulty, the factor scores would prove to be reliable across one or the other facet, and it would not be necessary to limit generalization to an occasion. However, such was not the case. An analysis of the estimates of variance components shown in Tables 49-53 shows that estimates for Age x subjects(coders) and Task x subjects(coders) were both substantial for all factors irrespective of the size of the estimate for the main effect of Age or Task.

For example, for Factor 1 the estimate of variance component for Age is 7.93 while the estimate for Task is -.18. Yet the estimate for As(c) is 8.42 and the estimate for Ts(c) is 11.33. For Factor 5 the estimate of variance component for Age is .79 and for Task is 45.96, while the estimate for As(c) is 18.36 and the estimate for Ts(c) is 26.29.

The generalizability coefficients were again estimates of reliability and are shown in Table 55. Only Factors 3 and 5 had values above .70 for the coefficient $\hat{\rho}^2(s,c,A^*,T^*)$. There are at least two possible explanations for the discrepancy in findings between this analysis and the previous analysis. First, this analysis included the 19 week session whereas the other analysis estimating reliability did not. It is possible that differences between coders at this earlier session were substantial and, therefore, the reliability was reduced accordingly. Unfortunately, there is not enough matched data at this session to test this possible explanation, but the intercoder agreement analysis at 25 weeks would seem to lend it some support.

A second explanation could be that since in this second design variance due to coders and variance due to groups to which coders were assigned was completely confounded, the inclusion of the variance due to coders as part of observed score variance (see Table 54) is making the estimate of reliability too conservative. Accordingly, the generalizability coefficients were computed with a different definition of observed score variance and the two sets of coefficients are shown in Table 56. The only notable difference is that the value of the coefficient $\hat{\rho}^2(s,c,A^*,T^*)$ for Factor 1 under method B is above .70.

Table 56

Comparison of Generalizability Coefficients for Reliability Analysis When Subjects Were Nested Within Coders

Generalizability Coefficient	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5	
	A[a]	B[b]	A	B	A	B	A	B	A	B
$\hat{\rho}^2(s,c,A,T)$	.20	.22	.21	.22	.30	.33	.20	.20	.24	.25
$\hat{\rho}^2(s,c,A^*,T)$	.39	.44	.38	.40	.52	.54	.38	.39	.44	.45
$\hat{\rho}^2(s,c,A,T^*)$	.46	.50	.45	.47	.58	.61	.41	.42	.53	.54
$\hat{\rho}^2(s,c,A^*,T^*)$	.66	.72	.63	.66	.78	.83	.59	.60	.73	.74

a Observed score variance under condition A = $\hat{\sigma}_{s(c)}^{2} + \frac{1}{O}\hat{\sigma}_{As(c)}^{2} + \frac{1}{W}\hat{\sigma}_{Ts(c)} + \frac{1}{OW}\hat{\sigma}_{residual}^{2} + \frac{1}{V}\hat{\sigma}_{c}^{2} + \frac{1}{VO}\hat{\sigma}_{cA}^{2} + \frac{1}{VW}\hat{\sigma}_{cT}^{2} + \frac{1}{VOW}\hat{\sigma}_{cAT}^{2}$

b Observed score variance under condition B = $\hat{\sigma}_{s(c)}^{2} + \frac{1}{O}\hat{\sigma}_{As(c)}^{2} + \frac{1}{W}\hat{\sigma}_{Ts(c)}^{2} + \frac{1}{OW}\hat{\sigma}_{residual}^{2}$

The coefficient $\hat{\rho}^2(s,c,A^*,T^*)$ in the split-plot design is comparable to the coefficient $\hat{\rho}^2(s,c,O^*)$ in the randomized block factorial design. Table 57 is a comparison of generalizability coefficients under the three conditions mentioned previously. Notice that for Factors 1, 2 and 5 the coefficient generated under the split-plot design is lower than that generated under the factorial design, whereas for Factors 3 and 4 the opposite is true. However, if .70 is used as the cutoff point in determining whether or not a factor is reliable, only the decision concerning Factor 1 would change under the A condition for the coefficient $\hat{\rho}^2(s,c,A^*,T^*)$ when compared to the coefficient $\hat{\rho}^2(s,c,O^*)$; if the B condition were used, the decisions made about the reliability of the factor scores would be exactly the same.

Analysis of Reliability

Medley and Mitzel (1963) identified two major reasons for the unreliability of data. The first was that separate measures of the same subject tend to differ too much across occasions. There were three major aspects of these differences: (1) items which enter into measurement lack consistency; (2) there is a lack of coder agreement; and (3) the behavior of the subject is unstable. The second major reason was that differences between subjects were too small. The following is an analysis of the five RCS factors according to these criteria.

Item consistency. There were several RCS behaviors with which we had difficulty during coder training which may be defined as "item consistency" problems. One was the item "01 - Baby Warms" (see Table 2). This was supposedly a measure of non-task related behavior in the young infant. Also, part of the definition for this behavior included

Table 57

Comparison of Generalizability Coefficients for Reliability Analysis When Generalization Was Limited To A Specific Occasion

Generalizability Coefficient	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
(s,c,O*)	.79	.69	.75	.52	.78
(s,c,A*,T*)[a]					
A	.66	.63	.78	.59	.73
B	.72	.66	.83	.60	.74

a See Table 56 for explanation of A and B

"self-reinforcing behavior" such as thumb sucking or putting a toy in the mouth. However, was the infant who waved a hand in front of his/her face and then put it in the mouth exhibiting "non-task related positive affect"? If the coder must decide whether that behavior is initiating (behavior with no task-related antecedent), or warms, it is easy to see why there would be lack of coder agreement. My recommendation would be to change the definition of this item to:

> Baby warms - task or non-task related behavior in which the infant smiles, laughs, gurgles, cooes, etc. (i.e., actively exhibits positive affect). This behavior may be seen at the same time as other infant behavior in which case, if there is parent verbal behavior, that parent behavior would be double coded. For example, if a baby reaches out and grabs a rattle while smiling at the same time in response to parent verbal behavior, the proper coding would be: 05 parent behavior 01 parent behavior.

This would eliminate the problem of defining what is "self-reinforcing" to the infant and would also eliminate the necessity of distinguishing between task- and non-task-related behavior.

A second item that caused some difficulty was "02 - Baby accepts" (see Table 2). The less-than-six-month-old infant would quite frequently look at what the parent was doing, glance away briefly, again look at the parent, again glance away, etc. If one tried to code every behavior change it became almost impossible to get any intracoder stability, let alone intercoder agreement. Therefore, I would add to the definition:

> Baby accepts - infant must be orienting visually to parent. If infant glances away briefly, but is spending the majority of the time period observing the parent, then record only one behavior.

A third problem that arose was the coding of baby babbling. It was very difficult to distinguish between baby sounds except in relation to other on-going behavior. Therefore, I would add an addendum to the

description of baby behavior (see Table 2):

> Baby babbling - (a) code as <u>baby responds</u> if the babbling follows a parent verbal behavior such as accepts, elicits, initiates, etc., (b) code as <u>baby initiates</u> if parent behavior does not immediately preceed, (c) code as <u>baby amplifies</u> if the baby sound builds upon (is noticeably different from) previous sounds.

A fourth problem was the coding of parent behaviors such as clicking, cooing, or other non-speech related sounds. They were quite obviously verbal behavior, but their meaning was not at all clear. My recommendation would be to code these as <u>parent initiates</u> although a case could also be made for coding these as <u>elicits</u> or <u>amplifies</u> depending upon the circumstances.

Overall I believe that the descriptions of behaviors for the RCS items were quite functional. For the vast majority of behaviors exhibited by both infant and parent it was not too difficult to come to some sort of agreement if the video-tape of a session was viewed by two or more persons at the same time and the observers were allowed to discuss how the behavior was to be coded. The major problem seemed to be inconsistent coding of the more subtle behaviors by one of the observers when coding alone.

<u>Coder Agreement</u>. A second part of the problem of reliability has to do with lack of coder agreement. The second analysis of intercoder agreement showed that the factor scores of parent-infant interaction at 37 weeks were quite acceptable. However, the coefficients at 25 and 43 weeks were not acceptable, with the exception of Factor 5 at 25 weeks. At 25 weeks it is possible that the ambiguity of the babies' behavior was a problem for all factors in addition to the lack of variance between subjects for Factors 1 and 3. At 43 weeks, however, differences between

coders were a major problem. This was not likely a result of a lack of training or ambiguity of behaviors since the intercoder agreement at 37 weeks was high. It was more likely a result of one of the coders attempting to code too quickly. Our experience indicated that it took about 50 minutes to code 10 minutes of tape and that two hours of coding at one time was about the maximum that one could code before becoming less accurate. One coder got behind because of a heavy class load and apparently was coding too quickly in a effort to catch up. The difficulty in keeping these types of coding errors under control is emphasized by the fact that most of the data for the 43 week session was coded in about three weeks.

It is my opinion that much of the coder disagreement was a result of two different problems. One was initial differences between coders, and the second was the instability of coder behavior. With respect to the first problem, one coder seemed absolutely compulsive about coding every single behavior while the other was much less so. Throughout the coding we had frequent sessions where the coders would first code a piece of tape separately and then code it together. Invariably, there would be differences which would be ironed out only to reappear at the next meeting. As mentioned above, the second problem was basically a result of external pressures on the coders, especially pressures centering around getting the job completed on time. Intracoder stability is not a problem that has been extensively discussed in this literature. Perhaps more work is needed in this area.

Stability of behavior. A third part of the problem of measures differing across occasions is that the behavior of the subject is unstable.

Another way of saying this is that the behavior either varies unsystematically or varies systematically across levels of variables for which we have not accounted. Medley and Mitzel (1963) and McGaw et al. (1972) both state that it is this third source of error variance which is the most important and the present study confirms their findings. Table 58 shows the sums of squares and the percent of total sums of squares (a measure of variance explained) for each source of variance for the five factors when subjects were crossed with coders. Notice that the main effect of coder accounts for one percent or less for each factor except Factor 3, where it accounts for three percent. The coder x block (a block being designated as a parent-infant pair) accounts for between four and eight percent. However, the main effect of Occasion accounts for over 25 percent with the exception of Factor 2, and the Occasion x block interaction accounts for between 21 and 43 percent.

Table 59 is a summary of the same information when subjects were nested within coders. Although the numbers are different because of the confounding of the variance due to coder and groups to which coders were assigned and the inclusion of an extra session in the split-plot analysis, the overall story is still the same. With the exception of Factor 4, the variance due to the main effect of coder is substantially less than one of the main effects of either Age of Infant or Type of Task. Comparing the results of the two analyses, it appears that the original differences between groups to which coders were assigned was influencing the variance due to coders such that differences were magnified, whereas for Factor 3 the influence is such that those differences were being concealed.

Table 58

Sums of Squares and Percent of Total Sums of Squares for Each Source of Variance for Factors of Parent-Infant Interaction When Subjects Were Crossed With Coders

	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5	
Source	SS	% of Total	SS	% of Total	SS	% of Total	SS	% of Total	SS	% of Total
Explained	337.79	29	385.10	17	527.81	37	1742.28	53	1120.76	27
coder	17.90	1	12.58	1	89.56	6	10.97	0	3.04	0
Occasion	295.47	26	339.94	14	435.03	31	1601.13	49	1107.28	27
cO	24.42	2	32.58	2	3.22	0	130.18	4	10.44	0
Unexplained	808.08	71	2023.36	83	906.14	63	1520.10	47	3055.95	73
blocks	334.60	29	510.29	21	319.30	22	374.53	11	775.20	19
cb	46.08	4	154.96	6	63.44	5	245.63	8	262.12	6
Ob	328.28	29	1035.59	43	446.52	31	669.41	21	1758.45	42
residual	99.12	9	322.52	13	76.88	5	230.53	7	260.18	6
Total	1145.87	100	2408.46	100	1433.95	100	3262.38	100	4176.71	100

Table 59

Sums of Squares and Percent of Total Sums of Squares for Sources of Variance for Factors of Parent-Infant Interaction When Subjects Were Nested Within Coders

	Factor 1		Factor 2		Factor 3		Factor 4		Factor 5	
Source	SS	% of Total	SS	% of Total	SS	% of Total	SS	% of Total	SS	% of Total
Explained	1942.14	33	2634.50	31	2073.61	38	5590.34	40	7559.90	46
coder	321.59	6	77.18	1	47.19	1	2214.93	16	43.31	0
Interactions involving coder	375.72	6	1029.07	12	316.94	5	729.85	5	875.32	6
Age of Infant	1032.79	18	129.77	2	1206.20	22	820.04	6	312.76	2
Type of Task	.01	0	832.66	10	79.91	1	734.35	5	5155.74	32
Other Interactions	212.03	3	565.82	6	423.37	8	1091.17	8	1171.77	8
Unexplained	3799.46	67	5861.31	69	3344.34	62	8361.98	60	8754.05	54
subjects(cI)	1091.42	19	842.12	10	1248.76	23	1597.76	11	2382.67	15
Interactions involving subjects(cI)	2708.04	48	5019.19	59	2095.58	39	6764.22	49	6371.38	39
Total	5741.60	100	8495.81	100	5417.95	100	13952.32	100	16312.95	100

Even though the facets used in each of the analyses explains some portion of the variance for each of the factors, there is still a large portion of variance which is not explained. It is possible that the behavior is truly "unstable" but there is also the possibility that some, as yet unidentified, facet can account for additional variance. One possibility might be a measure of infant temperament such as that developed by Thomas et al. (1963) and refined by Pederson, Anderson and Cain (1976). Another possibility might be a measure of parent sex-role as developed by Bem and her associates (1974, 1976). In any case there is still much parent-infant interaction variance as measured by these five factors to be explained over and above that accounted for by the five facets used in this study.

Differences between subjects. A second major reason cited by Medley and Mitzel (1963) for measures being unreliable was that differences between subjects are too small. In the intercoder agreement analysis this was cited as a possible cause of difficulty for Factors 1 and 3 at the early age grouping although overall it was not as significant a problem as one might expect when studying as restricted a group as we had. With respect to the two analyses of reliability it would appear that Factors 2 and 4 were the ones that were most affected by this problem. It was especially crucial for Factor 4 since the coder x block interaction in the analysis where subjects were crossed with coders accounted for almost as much variance as did blocks (8% and 11%, respectively); in the analysis where subjects were nested within coders the main effect for coders was larger than the main effect of blocks (16% and 11%, respectively).

Table 60 provides a summary judgment of sources of unreliability for each factor of parent-infant interaction. The subjects' behavior was unstable to the extent that the facets of coder and/or Occasion (or Infant Age Grouping and Type of Task) had to be considered before the generalizability coefficients computed for each facet in the reliability studies rose above .70. Observer disagreement was cited as a problem at both the 25 week and 43 week session according to the intercoder agreement analysis. The first reliabilty analysis (where subjects were crossed with coders) suggested that there may be some problem with coder disagreement for Factor 3. The second reliability analysis (where subjects were nested within coders) suggested there may be some problem for Factor 1. However, both reliability analyses suggested there were substantial problems with coder disagreement for Factor 4. Little variance between subjects was also cited as a possible problem for Factors 1 and 3 before 37 weeks and a possible problem for Factor 2 overall. Items lacking consistency was also cited as a problem for Factors 1, 3, and 4.

Additional Research

One of the main reasons for investigating the generalizability of measures is to provide some information about the reliability of measures prior to their being used in a decision study. One type of decision study that could be done would be to attempt to establish the predictive validity of these measures by studying their relationship with an infant competence measure such as the Bayley MDQ. Also, the investigation of the generalizability of these measures could be useful in designing future studies of parent-infant interation. As explained below, this

Table 60

Summary Judgment of Sources of Unreliability for Factors of Parent-Infant Interaction

Reason	Factor 1 Parent-Infant Interaction; Baby Amplifies	Factor 2 Parent-Infant Interaction; Baby Responds	Factor 3 General Baby Behavior	Factor 4 Parent-Infant Interaction; Baby Warms, Accepts	Factor 5 Parent-Infant Interaction; Baby Initiates
Behavior is unstable	Yes	Yes	Yes	Yes	Yes
Observer Disagreement	Substantial for intercoder agreement analysis at 25 weeks; moder- for second reliability analysis	Moderate for intercoder agreement analysis at 25 weeks and 43 weeks	Moderate for intercoder agreement analysis at 25 & 43 weeks; moderate for first reliability analysis	Substantial for intercoder agreement analysis at 25 weeks and both reliability analyses	Substantial at 43 weeks
Items lack consistency	May be a problem		May be a problem	May be a problem	
Lack of variance between subjects	May be a problem before 37 weeks	May be a problem overall	May be a problem before 37 weeks	No apparent problems	No apparent problems

study provides useful information for both of these endeavors.

It must be remembered that a generalizability coefficient is an estimate of the relationship of an observed score and a corresponding universe score. Generalization across all levels of facets represents a "larger universe" in that one can generalize a subject's score to any observation being considered. If there is little or no generalizability for a subject's score in this case then one is forced to look at a first-order coefficient of facet by subject in an attempt to increase the reliability. However, every higher order coefficient represents a "smaller universe" over which one can generalize. Therefore, it is better to pick the level of analysis in a decision study as low in the hierarchy as possible. Unfortunately, in the present case it does not seem feasible to use less than a first-order interaction of facet by subject. It was only after generalization was not intended to all levels of Occasion (or Infant Age Grouping and Type of Task) that generalizability coefficients over .70 were attained.

It was shown that the main effect of the facet of Parent Sex was not an important source of variance, but that the Parent Sex x block interaction was important for Factors 1 and 3. This would suggest that, even though the Parent Sex was eliminated in the additional analyses, it would be important to be specific as to the parent involved when using these two factors in a D study. Unfortunately, there was only sufficient data about the facet of Infant Sex to state that it did not appear to account for enough variance to warrant its inclusion in the model.

In terms of planning future studies, it seems appropriate to note the relative importance of Occasion in accounting for variance as

compared to either coder or Parent Sex in the design where coders were crossed with subjects. Using intercoder agreement as a measure of reliability simply does not consider this important source of variance. While only six subjects were used in this analysis compared with the 14 to 18 subjects used at each session in the intercoder agreement analysis it would appear that the reliability analysis provided additional valuable information. Notice that six subjects observed at three sessions by two observers is equivalent to 18 subjects observed at one session by two observers. Since resources for coding seem to always be less than one would prefer, it would seem preferable to observe a smaller number of subjects over several occasions rather than what has traditionally been done with intercoder agreement analysis.

It also seems appropriate to note the relative importance of Age of the Infant and the Type of Task in accounting for variance as compared with either Parent Sex or Infant Sex in the design where subjects were nested within coders. While it is true that this particular design is less powerful in showing differences with respect to Infant Sex than the other variables (Kirk, 1968), still this analysis emphasized similarities rather than differences in activity level and parent-infant interaction patterns for boys and girls during the first year of life. It also points to the powerful influence of situational variables in influencing the behavior of parents and infants and emphasizes the need to validate our findings in a wide variety of contexts, especially in the home. I would suggest four different types of tasks which might prove to be important: 1) Caretaking activities such as feeding, bathing, diapering, etc., 2) Freeplay with a set of toys appropriate for the age of the infant,

3) Structured teaching which involves an object, and 4) Structured teaching which does not involve an object, such as making sounds, clapping hands, singing songs, etc. It should be remembered that our data indicate a larger amount of score variance when the task is slightly difficult for the infant. Also, Llabre (1978) showed that by increasing the number of occasions to be sampled, reliability estimates could be increased dramatically.

Summary and Conclusions

This study investigated intercoder agreement and reliability of factor scores of parent-infant interaction through the use of generalizability theory. Three different designs were used: one for intercoder agreement at three different occasions and two for reliability. Each of the analyses was done a second time, eliminating the facets of Parent Sex and/or Infant Sex. The information derived from these three analyses led to the conclusion that even though the intercoder agreement coefficients were substantial at 37 weeks, generalization about a parent-infant pair's score could not be made from one occasion to another. It was also shown that the main effects of Age of Infant and Type of Task accounted for more variance than did either Sex of Parent or Sex of Infant.

With respect to the success of this particular study, it has been shown that information about the reliability of factors of parent-infant interaction using a split-plot design can approximate that derived from a randomized design, but because variance due to coders is confounded with variance due to group to which coder is assigned, the former cannot be substituted for the latter. However, since resources for coding are

usually limited, an investigator using a factorial design must use a small number of subjects, whereas with the split-plot design the entire sample of a study can be used. It therefore seems appropriate to view these two designs as complementary aspects of a reliability analysis rather than competitive. By doing so researchers can use systematic observation and report detailed information about the reliability of their data without going to the extra expense of double coding the entire sample. Consequently, there is clearly little excuse for the continued use of coefficients of intercoder agreement as measures of reliability.

GLOSSARY OF TERMS

GLOSSARY OF TERMS

ANOVA	Short-hand term for analysis of variance.
Decision study	A study from which decisions are made as to the relationship of the score under investigation to other measure (compare to generalizability study).
Facet	A source of score variability in addition to the between person variability (a factor or independent variable in ANOVA terminology with the exception that subject is never considered a facet).
Fixed variable	A variable in which all levels about which inferences are to be made are included in an experiment.
Generalizability coefficient	The ratio of the universe score variance to observed-score variance.
Generalizability study	A study which is done for the purpose of investigating the relationship between an observed score and a universe score (compare to decision study).
Intercoder agreement	The correlation between scores based on observations made by different coders on the same occasion.
Intraclass correlation coefficient	The ratio of two variances both of which are estimates of the same population.
Intracoder stability	The correlation between scores based on observations made by the same coder on different occasions.
Matched ratings	The case where every subject is rated by all coders.
MDQ	Bayley Mental Development Quotient: an assessment of infant perceptual-motor competence.
Observation composite or factor	A procedure for combining individual scores to assign composite or factor scores to each of the subjects; it is assumed that these scores reflect some characteristic of the behavior of that subject.
Observation instrument	A set of procedures whereby an observer can record and categorize the behavior of a subject or subjects. It normally consists of a number of items, to which the observer responds in some way dependent on the behavior that was observed.

Observation measure	A procedure for using an observation record to assign scores to each of the subjects of observation; each score so assigned is assumed to reflect some characteristic of that subject.
Observation record	A set of data (usually in the form of symbols) which describes the behavior of one or more subjects during one or more periods of observation.
Observed score	The score actually obtained when rating a subject.
Occasion	A single point in time when an observation is made.
Random variable	A variable in which a random sample of a population of levels about which inferences are to be made is included in an experiment.
Reliability (Classical model)	The extent to which a test is consistent in measuring whatever it does measure; dependability, stability, relative freedom from errors of measurement.
Reliability (from Medley and Mitzel, 1963)	The extent to which the average difference between two measurements independently obtained for the same subject (i. e., obtained on two separate occasions by two different observers) is smaller than the average difference between two measurements obtained for different subjects.
Reliability coefficient (Classical model)	Generally, the coefficient of correlation between two forms of a test, between scores on two administrations of the same test, or between halves of a test, properly corrected. The three coefficients measure somewhat different aspects of reliability but all are spoken of as reliability coefficients.
Reliability coefficient (from Medley and Mitzel, 1963)	The correlation between scores based on observations made by different observers on different occasions. They recommend using intraclass correlation coefficients derived through ANOVA.
Universe score	The average of the population of ratings that might be made for an individual.

APPENDIX

DESCRIPTION OF

STRUCTURED TEACHING ACTIVITIES

19 Weeks

Two-Way Stretch

This game's aim is to give your baby practice in controlling things around him by using his body.

Take the toy — a small telephone rattle with an elastic strip attached — and dangle it near the baby. Encourage him to reach and grab for it. Use such words as "get," "grab," and "catch" while you're playing together.

When he does grasp it, pull gently away so there's some stretch between you and him. Get into a push-pull game with him, saying, "Pull. You'll pull and I'll pull." Then gently release it and repeat. Try it so that he uses both hands.

Be sure it's fun and not teasing. Keep the toy so he can get it when he makes an effort. Remember that the underlying principle you want to convey to your baby is that it's worthwhile trying to do things, that an effort on his part can have gratifying results.

When he makes sounds of pleasure because he has grabbed it, respond to these sounds by repeating them. Enjoy his enjoyment. Please feel free to sit in one of the chairs or on the floor. If you have an infant seat, it's quite alright to use it. The most important thing is for you and your baby to be as comfortable and natural as possible.

25 Weeks

Mirror and Toy

The aims of this game are to help the baby become aware of his own appearance and to give him experiences in seeing objects reflected in a mirror.

Place your baby in your lap so that he is facing the same direction that you are. Hold a mirror so that he can see himself. Point to his reflection and say, "I see ______ (your baby's name)." "Where is _____?" "Find _______." "Look at _______."

Pick up the objects on the tray, one at a time, move them behind your baby's head so that he can see them in the mirror along with himself. Name the objects, telling something about the object such as, "This is a ball and it is round." Then say, "Where is the ball?" as you remove it from the mirror's reflection.

37 Weeks

Hide-and-Seek

For the very young child, out of sight is out of mind. Now he's ready to learn that things exist even when he can't see them.

Begin with a simple game using a toy and some soft covering material, such as a blanket. Attract your baby's attention to the toy and then partly hide it under the blanket so your baby can still see a part of it.

Then say, "Where did it go?" "Find the Toy."

If he's puzzled and doesn't seem to know how to retrieve it, show him how. If he ignores the toy after it is hidden, play with it by yourself in front of him, but don't demand his attention or any action. He will, on his own, get interested in what you are doing.

Partly hide it again until he's able to get it himself.

Play the same game, but hide the toy completely under the soft material so he can see that something is under the blanket. Encourage him to lift it up and get his toy.

Repeat this for fun a number of times and then leave the child with both toy and blanket.

43 Weeks

Blocks

Since your child can now handle small objects with his fingers rather well, he's ready for block play. Blocks are perhaps the best of all possible toys because he can do so many things with them. Start him out with just a few.

Place two blocks in front of him while you're both sitting on the floor and show him how one can be put on top of the other. Let him do it. Then add a third so he can build a simple three-block tower. Don't worry if they're not directly one on top of the other. This is a self-correcting activity. If he doesn't build well enough, it will just tumble down. He will enjoy the tumbling as much as the building.

A variation of this is to show him how you can place two or three blocks in a line on the floor and push them around. If he pushes on the third one, the first two will go straight for a few seconds and then get out of line. He'll enjoy watching this happen, and gradually he'll gain the skill needed both to build the tower straight and keep the blocks in line.

You can also make up your own variations of block play. The main idea is to encourage him to develop his new found skills and for the parents and child to enjoy playing together.

REFERENCES

Ad Hoc Committee on Child Mental Health. Report to the Director, National Institute of Mental Health. Washington, D. C.: U. S. Department of Health Education and Welfare, National Institute of Mental Health, February, 1971.

Ainsworth, M. D. The development of infant-mother attachment. In B. M. Caldwell & H. N. Ricciuti (Eds.), Review of child development research III. Chicago: University of Chicago Press, 1973.

Ainsworth, M. D., Salter, S. M. Bell, & Stayton, D. J. Individual differences in the development of some attachment behaviors. Merrill-Palmer Quarterly, 1972, 18, 123-143.

Algina, J. Comment on Bartko's "On various intraclass correlation reliability coefficients." Psychological Bulletin, 1978, 85, 135-138.

Bakow, H., Sameroff, A., Kelly, P., & Zax, M. Relation between newborn behavior and mother-child interaction at 4 months. Paper presented at the meeting of the Society for Research in Child Development, Philadelphia, March, 1973.

Bales, R. F. Interaction process analysis. Cambridge, Mass.: Addison-Wesley, 1951.

Bayley, N. Bayley scales of infant development. New York: The Psychological Corporation, 1969.

Bell, R. Q., & Harper, L. Effects of the young on adults in man and other animals. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1977.

Bell, S. M., & Ainsworth, M. D. S. Infant crying and maternal responsiveness. Child Development, 1972, 43, 1171-1190.

Bem, S. L. The measurement of psychological androgyny. Journal of Consulting and Clinical Psychology, 1974, 42, 155-162.

Bem, S. L., Mahtyna, W., & Watson, C. Sex typing and androgyny: Further explorations of the expressive domain. Journal of Personality and Social Psychology, 1976, 34, 1016-1023.

Biller, H. B. The father-infant relationship: Some naturalistic observations. Unpublished manuscript, University of Rhode Island, 1974.

Bowlby, J. Maternal care and mental health. Geneva: Who, 1951.

Bowlby, J. Attachment and loss (Vol. 1) Attachment. New York: Basic Books, 1969.

Boyer, E., Simon A., & Karafin, G. Measures of maturation: An anthology of early childhood observation instruments, 3 volumes. Philadelphia: Research for Better Schools, 1973.

Brennan, R. L. The calculation of reliability from a split-plot factorial design. Educational and Psychological Measurement, 1975, 35, 779-788.

Brody, S. Patterns of mothering. New York: International University Press, 1956.

Bromwich, R. M. Focus of maternal behavior in infant intervention. American Journal of Orthopsychiatry, 1976, 46, 439-446.

Caldwell, B. M., & Honig, A. S. APPROACH: A procedure for patterning responses of adults and children: Coding manual. Little Rock, Arkansas: University of Arkansas, Center for Early Development and Education, 1970.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963 (Reprinted from N. L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally, 1963).

Clark-Stewart, K. A. Interactions between mothers and their children: Characteristics and consequenses. Monographs of the Society for Research in Child Development, 1973, 38, (6-7. Serial No. 153).

Crawley, S. B., Rogers, P. P., Freidman, S., Iacobbo, M., Criticos, A., Richardson, L., & Thompson, M. A. Developmental changes in the structure of mother-infant play. Developmental Psychology, 1978, 14, 30-36.

Cronbach, L. J., Gleser, G. C., Nanda, J.Z., Rajaratnam, H. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Cronbach, L. J., Rajaratham, N., & Gleser, G. C. Theory of generalization: A liberalization of reliability theory. The British Journal of Statistical Psychology, 1963, 16, 137-163.

Dixon, W. J. (Ed.). BMD: Biomedical computer programs. Berkeley, California: University of California Press, 1974.

Emde, R. N., Gaensbauer, T. J., & Harmon, R. J. Emotional expression in infancy: A biobehavioral study. Psychological Issues, Monograph Series, 1976, 40, 1-118.

Escalona, S., & Corman, H. Early life experiences and the development of competence. International Review of Psycho-Analysis, 1974, 1, 151-168.

Flanders, N. A. Teacher influence, pupil attitude, and achievement. U. S. Department of Health, Education, and Welfare, Office of Education, Cooperative Research, Monograph No. 12. Washington, D. C.: Government Printing Office, 1965.

Gewirtz, H. & Gerwirtz, J. Visiting and caretaking patterns for kibbutz infants: Age and sex trends. American Journal of Orthopsychiatry, 1968, 38, 427-443.

Gewirtz, J. Levels of conceptual analysis in environment-infant interaction research. Merrill-Palmer Quarterly, 1969, 15, 7-48.

Goldberg, S. & Lewis, M. Play behavior in the year-old infant: Early sex differences. Child Development, 1969, 40, 21-31.

Gordon, I. J. Baby learning through baby play. New York: St. Martin's Press, Inc., 1970.

Gordon, I. J. An investigation of the social roots of competence. Gainesville, Florida: University of Florida, Institute for Development of Human Resources, final report to NIMH on Project No. 1, R01 MH 22724, October 1974.

Gordon, I. J. Parenting, teaching, and child development. Speech delivered at the Mid-West Association for the Education of Young Children, Madison, Wisconsin, April, 1975.

Gordon, I. J. & Jester, R. E. Instructional strategies in infant stimulation. JSAS Selected Documents in Psychology, 1972, 2, 122.

Gordon, I. J. & Soar, R. S. Parent-infant transactions and infant competence. University of Florida: Institute for Development of Human Resources. National Institute of Mental Health Grant No. 1 R01 MHIHD 27480-01, In progress, begun in 1976.

Hollingshead, A. B. Two-factor index of social position. Unpublished manuscript, 1957.

Horn, J. L. An empirical comparison of methods for estimating factor scores. Educational and Psychological Measurement, 1965, 25, 313-322.

Kagan, J. Change and continuity in infancy. New York: John Wiley & Sons, 1971.

Kane, M. T. & Brennan, R. L. The generalizability of class means. Review of Educational Research, 1977, 47, 267-292.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, California: Brooks/Cole Publishing Co., 1968.

Kogan, K. L., Wimberger, H. C. & Bobbitt, R. A. Analysis of mother-child interaction in young mental retardates. Child Development, 1969, 40, 799-812.

Kotelchuck, M. The infant's relationship to the father: Experimental evidence. In M. E. Lamb (Ed.). The role of the father in child developement. New York: John Wiley & Sons, 1976.

Lamb, M. E. Fathers: Forgotton contributors to child development. Human Development, 1975, 18, 245-266.

Lamb, M. E. Effects of stress and cohort on mother- and father-infant interaction. Developmental Psychology, 1976a, 12, 435-443.

Lamb, M. E. Interactions between eight-month-old children and their fathers and mothers. In M. E. Lamb (Ed.). The role of the father in child development. New York: John Wiley & Sons, 1976b.

Lamb, M. E. The role of the father: An overview. In M. E. Lamb (Ed.). The role of the father in child development. New York: John Wiley & Sons, 1976c.

Lamb, M. E. Twelve-month-olds and their parents: Interaction in a laboratory playroom. Developmental Psychology, 1976d, 12, 237-244.

Lamb, M. E. Father-infant and mother-infant interaction in the first year of life. Child Development, 1977, 48, 167-181.

Lewis, M. A. The meaning of a response, or why researchers in infant behavior should be oriental metaphysicians. Merrill-Palmer Quarterly, 1967, 13, 7-18.

Lewis, M. A. & Goldberg, S. Perceptual-cognitive development in infancy: A generalized expectancy model as a function of mother-infant interaction. Merrill-Palmer Quarterly, 1969, 15, 81-100.

Llabre, M. M. An application of generalizability theory to the assessment of writing ability. Doctoral dissertation, University of Florida, 1978.

Lusk, D. & Lewis, M. Mother-infant interaction and infant development among the Wolof of Senegal. Human Development, 1972, 15, 58-69.

Lytton, H. Observation studies of parent-child interaction: A methodological review. Child Development, 1971, 42, 651-684.

McGaw, B., Wardrop, J. L. & Bunda, M. A. Classroom observation schemes: Where are the errors? American Educational Research Journal, 1972, 9, 13-27.

Medley, D. M. & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.). Handbook of research on teaching. Chicago: Rand McNally, 1963, 247-328.

Moss, H. A. Sex, age, and state as determinants of mother-infant interaction. Merrill-Palmer Quarterly, 1967, 13, 19-36.

Ober, R. L., Wood, S. E., & Roberts, A. An instructional manual for the Reciprocal Category System. Based on a paper presented at the annual meeting of the American Educational Research Association, Chicago, 1968.

Pederson, F. A., Anderson, B. J., & Cain, R. L., Jr. A methodology for assessing parental perception of infant temperament. Paper presented at the Fourth Biennial Southeastern Conference on Human Development, April, 1976.

Peterson, M. A. The study of a methodology for observing mother-infant interaction patterns during the second year of life. Dissertation Abstracts International, 1975, 36 (12-8), 895.

Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. Generalizability of stratified-parallel tests. Psychometrika, 1965, 30, 39-56.

Rebelsky, F., & Hanks, C. Fathers' verbal interaction with infants in the first three months of life. Child Development, 1971, 42, 63-68.

Rowley, G. R. The reliability of observational measures. American Educational Research Journal, 1976, 13, 51-59.

Sparling, J. J., & Gallagher, J. J. (Eds.). Research directions for the 70's in child development. Chapel Hill, North Carolina: University of North Carolina, 1971.

Spearman, C. The proof and measurement of association between two things. American Journal of Psychology, 1904, 15, 72-101.

Stern, D. Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis and L. A. Rosenblum (Eds.), The effects of the infant of its caregivers. New York: John Wiley and Sons, 1974.

Thomas, A., Chess, S., Birch, H. G., Hertzig, M. E. & Korn, S. Behavioral individuality in childhood. New York: New York University Press, 1963.

Walters, J., & Stinnett, N. Parent-child relationships: A decade review of research. Journal of Marriage and the Family, 1971, 33, 70-111.

Watson, J. S. Smiling, cooing, and 'the Game'. Merrill-Palmer Quarterly, 1972, 18, 323-339.

Watts, J. C., & Barnett, J. C. Environments compared. Unpublished manuscript, Harvard University, 1971.

Yarrow, M. L. Problems of methods in parent-child research. Child Development, 1963, 34, 215-226.

BIOGRAPHICAL SKETCH

William G. Huitt was born on June 8, 1944, in Cincinatti, Ohio. After spending his childhood years in Oregon and Texas, he moved to Frankfurt, Germany, where he graduated from high school in 1962.

Upon his return, Mr. Huitt enrolled in the University of Alabama extension center in Mobile. During the next seven years he both worked and went to school, graduating from the University of South Alabama in June, 1969, with a major in business administration and marketing. During this time period he joined a medical unit of the Alabama National Guard and spent six months on active duty. He married Marsha K. Anderson in March, 1968.

Mr. Huitt was awarded an Educational Research Training Fellowship at the University of Florida and began graduate school in September, 1969. He obtained an M.Ed. degree in 1971, majoring in business education and educational research. He taught eighth-grade mathematics for one year and then taught in a Business Administration Program at Albuquerque Technical-Vocational Institute in Albuquerque, New Mexico.

In June, 1973, Mr. Huitt and his wife joined the Peace Corps and went to Truk District, Micronesia. He was assigned to a fishing cooperative as a marketing and management advisor and taught at the teacher training center. He also served as a statistics and research consultant to the Truk Education Department.

In June, 1975, he returned to graduate school in educational psychology and was awarded an assistantship with the Parent Education

team of Project Follow Through. In December, 1975, he was blessed with the birth of twin boys, Kevin and Geoffrey, and shortly thereafter began work with Dr. Michael Resnick as a parent educator. In July, 1976, he became project director of the Parent-Infant Transaction Project under the direction of Dr. Ira J. Gordon. Mr. Huitt is currently employed as a Research and Evaluation Specialist with Research for Better Schools, Inc., in Philadelphia, Pennsylvania.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Barry J. Guinagh

Dr. Barry J. Guinagh, Chairman
Associate Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert Soar

Dr. Robert S. Soar, Cochairman
Professor of Foundations of Education

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Robert C. Ziller

Dr. Robert C. Ziller
Professor of Psychology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

William B. Ware

Dr. William B. Ware
Professor of Foundations of Education

This dissertation was submitted to the Graduate Faculty of the Department of Foundations of Education in the College of Education and to the Graduate Council, and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December, 1978

Robert R. Sherman

Chairman, Foundations of Education

Dean, Graduate School